POZNAŃ UNIVERSITY
OF ECONOMICS
AND BUSINESS

Izabella Krzemińska

Automatic Data-Based Personality Assessment as a Method of Electronic Services Auto-Personalisation

Automatyczne wyznaczanie osobowości z danych jako metoda auto-personalizacji usług elektronicznych

Doctoral Dissertation

Supervisor:

Prof. Dr. Witold Abramowicz

Auxiliary supervisor: Dr. Marcin Szmydt

Date of submission:

Poznań 2022

Acknowledgments

Writing this dissertation, which is essential to me personally, was possible thanks to the support and help of many people, whom I would like to thank very much.

First of all, I am incredibly grateful to my supervisor, prof. Witold Abramowicz for his constant support, patience, vast knowledge and impressive experience, which he shared, leaving me with complete autonomy in my research.

I would also like to express my gratitude to the assistant supervisor, Dr. Marcin Szmydt, whose kindness and practical advice helped me overcome all obstacles in achieving my goal.

I would like to thank all lecturers and colleagues on the Data Science in Business Administration doctoral studies. Their kindness, support, and the built atmosphere at the Institute made the last 4 years a great time.

I also very appreciate all the support I received from my superiors and colleagues from Orange Labs Polska and Orange Labs in France. Without their approval of the purposes of this research, funding of data collection, and technical and application development support, the implementation of this significant project would not be possible.

Finally, I would like to express my gratitude to my closest family, husband Arthur, and my wonderful daughters, Aleksandra and Magdalena, for their love, spiritual support and their patient understanding of all the sacrifices and inconveniences because of my research passion.

Contents

Abbreviations

- **A** Agreeableness
- **ADA** AdaBoost Classifier algorithm
- **AI** Artificial Intelligence
- **API REST** Application Programming Interface Representational State Transfer
- **BC** Bagging Classifier algorithm
- **BI** Business Intelligence
- **BIG 5** Big Five (Personality Factors)
- **C** Conscientiousness
- **CART** Classification and Regression Trees algorithm
- **CDR** Call Detail Record
- **CRM** Customer Relationship Management
- **CV** Cross Validation
- **DT** Decision Tree algorithm
- **E** Extraversion
- **ERCC** Ensemble of Regression Chain Corrected
- **ETC** Extra Tree Classifier algorithm
- **FEDS** Framework for Evaluation in Design Science
- **FN** False Negatives
- **FP** False Positives
- **FPR** False Positive Rate
- **GDPR** General Data Protection Regulation
- **GUI** Graphical User Interface
- **HCE** Host Card Emulation
- **HRI** Human-Robot Interaction
- **HAI** Human-AI Interaction
- **IT** Information Technology
- **KNN** K-Nearest Neighbour algorithm
- **LDA** Linear Discriminant Analysis
- **LIWC** Linguistic Word Count
- **LR** Logistic Regression
- **LGBM** LGBM Classifier

- **MBTI** Myers-Briggs Type Indicator
- **ML** Machine Learning
- **MNB** Multi-modal Naive Bays algorithm
- **N** Neuroticism
- **NB** Naive Bays algorithm
- **NN** Neural Network
- **O** Openness to Experience
- **PAPA** Personality based Automatic personalisation of App method
- **PI** Permutation Importance Evaluation
- **RFECV** Recursive Feature Elimination With Cross-Validation
- **RF** Random Forest
- **RFFI** Random Forest Feature Importance
- **RPCA** Random Principal Component Analyses
- **S** Emotional Stability
- **SDG** Sustainable Development Goals
- **SGD** Stochastic Gradient Descents
- **SIP** System Initiated personalisation
- **SML** Supervised Machine Learning
- **SMOTE** Synthetic Minority Over-sampling Technique
- **S-SML** Semi-Supervised Machine Learning
- **SVC** Support Vector Classification
- **SVM** Support Vector Machine algorithm
- **TN** True Negatives
- **TQM** Total Quality Management
- **TP** True Positives
- **UIP** User-Initiated personalisation
- **UISPP** User's Initial Smartphone Personality Profile
- **USML** Un-Supervised Machine Learning
- **UX** User Experience
- **XGB** XG Boost Classifier

Chapter 1

Introduction

1.1 Motivation

We operate in times of constant digital transformation. Habits, ways of performing tasks and general human functioning are gradually enriched with many helpful digital tools. The low cost of electronics production, its constant minimisation, increasing the energy ergonomics of electronic devices, and finally, customers' growing needs and expectations make the digital transformation accelerate exponentially and become commonplace. It is not easy to imagine today's world without digital devices. What was impossible ten years ago is now available. The post-Covid-19 world and ubiquitous 'remote' functioning accelerated these processes of digitisation of societies and individuals. We are changing habits, ways of performing tasks and massively using personal digital devices such as smartphones. A new phobia of lack of access to a smartphone - nomophobia has already been identified by researchers Bianchi and Phillips (2005) and Yildirim and Correia (2015). It is a significant sign of ongoing digitisation and its impact on human lives. This kind of technological anxiety is tangible evidence of the revolutionary and even evolutionary changes that continue to happen nowadays.

However, despite the universality in using devices themselves, in the case of advanced digital services (e.g. smart home or virtual assistant/agent) there are still barriers related to the fear of artificial intelligence (AI), both in the physical (robot) and purely virtual (application, program, virtual entity, avatar). In the context of the ongoing digital transformation, traces of users' digital activity are more and more used to collect information about the client, profiling and classifying (Kosinski et al., 2012;

Kosinski et al., 2013). Good determinants are sought to diversify the needs of users in order to overcome these barriers actively. In today's world, old profiling methods are failing. Years ago, in 1997, Philip Kotler proposed segmenting consumer markets according to geographic, demographic, psychographic and behavioural variables. Psychographic segmentation of clients consisted of dividing them into groups based on lifestyle and personality (Kotler and Turner, 1997). Such segmentation techniques based on the selection of appropriate descriptors increased the effectiveness of marketing activities (Wind, 1978).

Additionally, the variables used for segmentation should always be considered in terms of their measurability, availability, reliability and the ability to discover the characteristics of each customer segment. Therefore, companies producing various consumer goods are looking for opportunities to introduce unique features to the marketing segmentation of customers (Cody, 2012). Researchers have previously attempted to segment clients based on their personality (Becker and Connor, 1981) because they found it valuable to recognise a deeper level of individual differences. They analysed the markets for various social, economic and personal characteristics. It turned out that clients from the same group, with respect to their demographic characteristics, may show very different psychographic profiles (Kotler and Turner, 1997). This argues for the use of psychography for segmentation.

Before that, however, the problem was to obtain a psychographic description on a massive scale. Before the ubiquitous digitisation of behaviour and events in real-time, psychography was implemented based on paper questionnaires. However, paper psychometric tools have many limitations, such as time-consumption, effort required, and factors that distort the results such as conscious or unconscious self-deception. All of this can be considered a disadvantage when used on a large scale for research and business purposes. Profiling based on readily available demographic characteristics (such as gender, age, place of residence, education, economic status) is commonly used when it is necessary to understand differences between groups. Profiling based on demographic characteristics can be successfully used in health research and modelling (Schäfer et al., 2012). However, when examining personal motives, fears, psychological barriers and drivers, there is a need for more detailed descriptions of human nature.

Additionally, in modern Industry 4.0, a deeper level of personalisation is required.

Nowadays, companies try to learn the requirements of unique individuals, not the requirements of a typical user (Kowalski and Weresa, 2019; Trzcieliński, 2020). The possibilities offered by the digital world in terms of analysing individual differences between people are constantly growing. This is mainly due to the Big Data era and ubiquitous collecting of the traces of digital activity, even if they are not currently required or used. However, the collection of event logs relating to the history of service usage applied to create the profiles usually require a lot of time, and resources. Therefore, the first motivation to write this dissertation is to look for good classifiers of behaviour and customer needs. Classifiers that will be calculated automatically can be used from the first moment of using the service. An additional motivation is to verify the possibility of using this knowledge to actively adjust the AI-based service to the user and his needs. With the high availability of digital data and the constantly growing computing power of the simplest devices, scientists have a much greater capacity than ever to study human behaviour.

Moreover, digitisation allows the study of behavioral traces, not laboratory behaviour or behaviour declarations, as was the case with surveys and polls. Such an environment has proven to be helpful for the rapid development of robotics. For the telecommunications operator (and all digital interactive service providers), it is crucial to lower the barriers related to the fear of new intelligent technologies based on artificial intelligence (AI) solutions and the increasing use of more advanced virtual agents.

Thinking about the use of customer digital data became popularized for business purposes as the Internet and social networking sites rapidly developed. Currently, the main driving force is the more widespread implementation of artificial intelligence in the construction of solutions for end-users and the continuous enrichment of the functionality of mobile phones. As cell phones have become an integral part of modern human life, they have access to more and more of our activities. Consequently, the search for determinants of the differentiation of user needs is increasingly based on data from smartphones. However, for fifteen years, the popularity of researching the potential of digital footprint and enriching the client's description with more detailed dimensions, such as personality, is observed. Ubiquitous digitisation has a rich research potential both in the area of personality diagnostics and service profiling. Personality, apart from socio-demographic characteristics and personal experiences that are difficult to verify, seems to be the essential element determining users' needs (this is confirmed also by the preliminary research for this dissertation - Appendix A).

There is also a need for faster and more individualised user classification and profiling, not based on a long history of events in the account (which is not available in the case of the new customer of the service), but based on a set of a small amount of information available from the start of using the service. Classifications based on the user's personality may even affect the shape of the service presented, like for example graphical user interface (GUI) or even configuration of service functionalities.

Since 2012, a significant increase in interest in profiling based on digital data has happened. Michał Kosiński's research on Facebook and Twitter Kosinski et al. (2012), and IBM Watson's algorithm widely known and even the Cambridge Analytica scandal was not able to stop this trend (Hinds et al., 2020). However it can be seen that personalisation algorithm is usually limited to data from the service and makes data available only to the service. This limitation is primarily due to the legal provisions on the protection of privacy. Open access to user data has fallen significantly after the Cambridge Analytica scandal. Deep machine learning techniques usually requires a large amount of data to build an algorithm and are based mainly on user activity logs from the service.

In November 2014, Amazon launched its first voice assistant, Alexa. The name is not accidental. The hard consonant with 'x' helps to be recognised with greater precision by Machine Learning (ML) speech recognition modules. The product became famous worldwide. Even though, initially, it could only communicate in English and French. According to Wikipedia sources [1] in June 2015, it had about 1000 functionalities. By 2018, there were already 45,000, and in April 2019 there were over 90,000 functionalities available to users.

Nevertheless, despite many such functionalities, customers limit their use only to those known from Google search engines, such as checking the weather, news or playing music from Spotify. Adjusting this type of service is time-consuming and limited by the cognitive abilities of users. In new, interactive and technologically advanced services based on artificial intelligence, the risk of rejection and discontinuation Alexa appears critical. Based on the Amazon Alexa experience, whose main purpose was to help buy

[1] source:(https://en.wikipedia.org/wiki/Amazon_Alexa)

on Amazon, 90% used it only once and never repeated the experiment (Anand, 2009 (accessed August 30, 2019)). The main reason is that people very rarely go beyond what they know. There is a risk that most of the service's 90,000 features will never be detected in such a situation. Therefore, user categorisation and service profiling should be performed as soon as possible after the service is installed. In this way, it will be possible to adapt the service to the user from the beginning and offer functionalities dedicated to him/her. Figure 1.1 shows a graph from the report Anand (2009 (accessed August 30, 2019)) with the most commonly used Amazon Alexa features.

Analysing the example of Amazon again (Fig.1.1), it can be said that users limit the service to basic functionalities known from personal experience. Because based only on experience, it is difficult to discover unknown functionalities available on the site, such as buying products. In this situation, usage history algorithms, as recommended, are not sufficient. According to the article, Amazon recognised that so few people buy things via Alexa and it was a big challenge for the company. Furthermore, the indicated direction of product development was to enrich it with mechanisms that diagnose and look for features that distinguish individual customers.

It should be noted that recommendations based on the analysis of behaviour in similar groups described by socio-demographic variables or based on segmentation will not map unused functionalities of the service. Additionally, a user must use the service to implement these practices. All these practices can be applied while using the services and are useless in case of abandonment of the service, discouragement or boredom. One possible solution is to have a basic user profile at the time of service installation and personalise the service from the first use according to the user's profile needs.

So the main problem with advanced smart services is finding a way to present the most appropriate functionalities for the user to minimise the risk of service dissatisfaction or rejection. It seems that a personality profile can be a helpful classifier of user needs when customising a service or application. Hence the direction of research proposed in the dissertation.

After the Cambridge Analytic scandal, an additional incentive was protecting data privacy in the personalisation process. Increasingly, the discussion of the use of digital data raises ethical issues. New economic mechanisms are described that use observations from the operation of digital giants, whose power is based on the re-use of digital

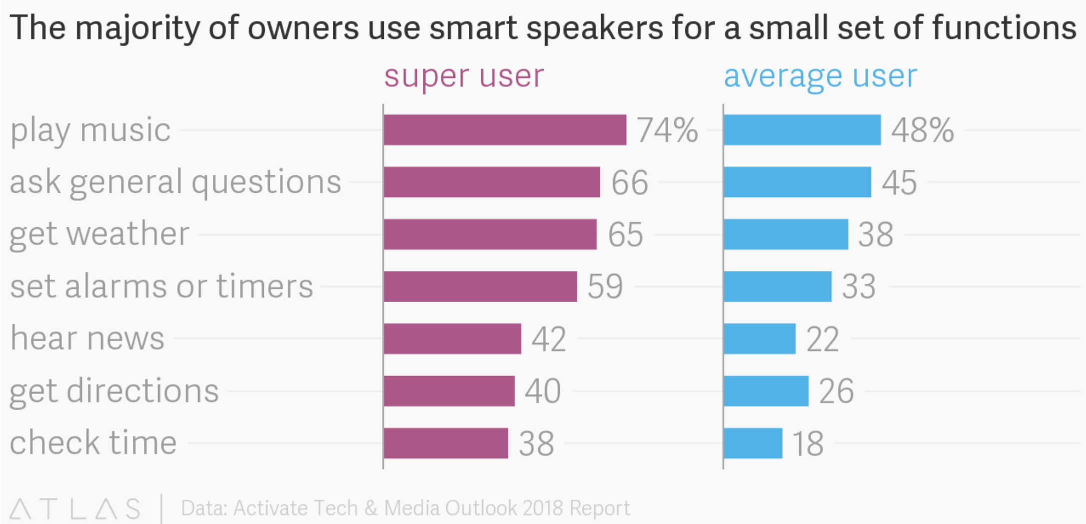


Figure 1.1. The most popular question to smart speaker in group of super users (min. 3 times a day) and average service users. Source: (Anand, 2009 (accessed August 30, 2019)).

data, most often sensitive data. Zuboff (2019) described a new type of economic system centred around the commodification of personal data in her book and called it surveillance capitalism.

The concept of generating profit from personal data to more accurately reach consumers was created in the Google company, specifically in Google AdWords. According to the author, earlier eras in industrial capitalism used nature and the environment, and surveillance capitalism applies totalitarian mechanisms to human nature. Undeniably, data collection can benefit individuals in society and society through optimisations such as smart cities. Using this data in the context of profit-making (advertising) and constantly changing purchasing behaviour (recommendations) may threaten human freedom, autonomy and respect for privacy.

The benefits for society (security, smart cities, making life easier) do not always exceed the costs for individuals. It is not easy to estimate how many users do not know that they are paying for these benefits with their private data and the proportion of the population that is subject to such data collection. In the preliminary qualitative research conducted for this dissertation, some respondents rationalised their decisions regarding the free sharing of their data on Google or Facebook. In 2016, CBOS conducted a survey in which questions were asked about the acceptance of access to digital data by state services such as the police or internal security services. 46% accepted this type of control to some extent, 30% were against, and 24% did not have an opinion or were ambivalent (Feliksiak, 2016).

Additionally, acceptance decreased with age and increased with the frequency of Internet use. The report of Amnesty International also drew attention to the problem of threats related to the use of digital data (Amnesty_International, 2019). Their report mentions specific platforms that collect data about users, such as Google and Facebook, based on algorithmic systems that process vast amounts of data to influence users' internet use. The Cambridge Analytica scandal showed how easy it is to use people's data in unpredictable ways to manipulate and influence them (Isaak and Hanna, 2018; Tufekci, 2018). The very existence of mechanisms and logic of action, called computational governance (one of the mechanism of surveillance capitalism), undermines human self-determination's psychological and political foundations (Amnesty_International, 2019). There is even the term 'informational self-determination', meaning the individual's right to decide when and at what restrictions, information about their private life should be passed on to others (Quinn, 2021). Legal regulations do not keep up with companies' ingenuity in obtaining, using and reselling data about users of electronic services. Most of these mechanisms are incomprehensible to an average digital service user.

In addition, voluntary consent to using behaviourally automated data as a necessary element of any electronic service no longer constitutes grounds for concluding that people knowingly make choices about sharing their data. Therefore, one of the reasons for conducting this study was to check the possibility of personalising the service while maintaining complete confidentiality of data and the profile itself. Personalisation can benefit the user and help him compensate for some personal imperfections and deficiencies, which does not necessarily mean losing user privacy.

To sum up, the proposed research was based on: both the desire to find classifiers of behaviours better differentiating human behaviour and motivations, and the wish to check the possibilities for their automatic detection from digital data. Such a process should be carried out with complete protection of privacy. Thus, both the data and the knowledge about the user, i.e. the profile, would remain private information accessible only to the service user, and not transmitted outside the terminal device.

An additional aim of the study was to check the possibility of using the knowledge about the user thus created for the automatic personalisation of the service. Under the assumption of complete privacy protection, this would mean that the mechanism for customising the service itself would have to be embedded in the service itself and would have to be fully automated.

1.2 Goal and Research Questions

The above motivations lead to the formulation of the following research questions.:

- (RQ.1) Is it possible to create an automatic method for determining the user's personality based on the mobile phone data available at the time of the application installation?
- (RQ.2) Is it possible to create an automatic personalisation with full privacy of users data (without the necessity of sharing data with the supplier)
- (RQ.3) Is it possible to use an automatic personality detecting method for personalising electronic services?

The primary objective of this doctoral dissertation is to develop novel, effective and accurate methods of assessing user's needs defined by the personality profile, that is, automatically indicated and, based on the data available from the moment of the service installation without the delay connected with collecting user's data log. The specific objectives of the work are:

(O.1) to develop the method for determining user's personality based on the limited amount of data available in the moment of service installation.
(O.2) to implement the created method (O.1) into a mobile application
(O.3) to assess the applied classifier of users needs by means of an experiment based on laboratory measurement or measurement on a test group

Thesis

Automatic identification of the user's personality profile, based on the data available during the installation of the electronic service, can be used as a classifier facilitating the personalisation of this service. The thesis is verified on an example of a class of electronic services; an application installed on smartphones.

1.3 Research Methodology

1.3.1 Design Science Research

The overall methodology applied within the research is Design Science based on the principles and framework proposed by A. Hevner and Chateerjee (2012). The main goal of the research in the field of design science is to gain knowledge. This knowledge is used by specialists in a given field to develop solutions to problems. The main paradigm proposed by Hevner is to gain knowledge and understanding of the problem domain by building and evaluating artefacts. During the research process, a designed artefact is evaluated through the practical application (A. Hevner and Chatterjee, 2010; A. R. Hevner et al., 2004). This methodology was chosen due to the cognitive and exploratory character of research on personality determination as well as the different nature of the goal that is evaluation of the quality of the artefact and not the determination of cause-effect relationships. Characterisation of the defined research problem

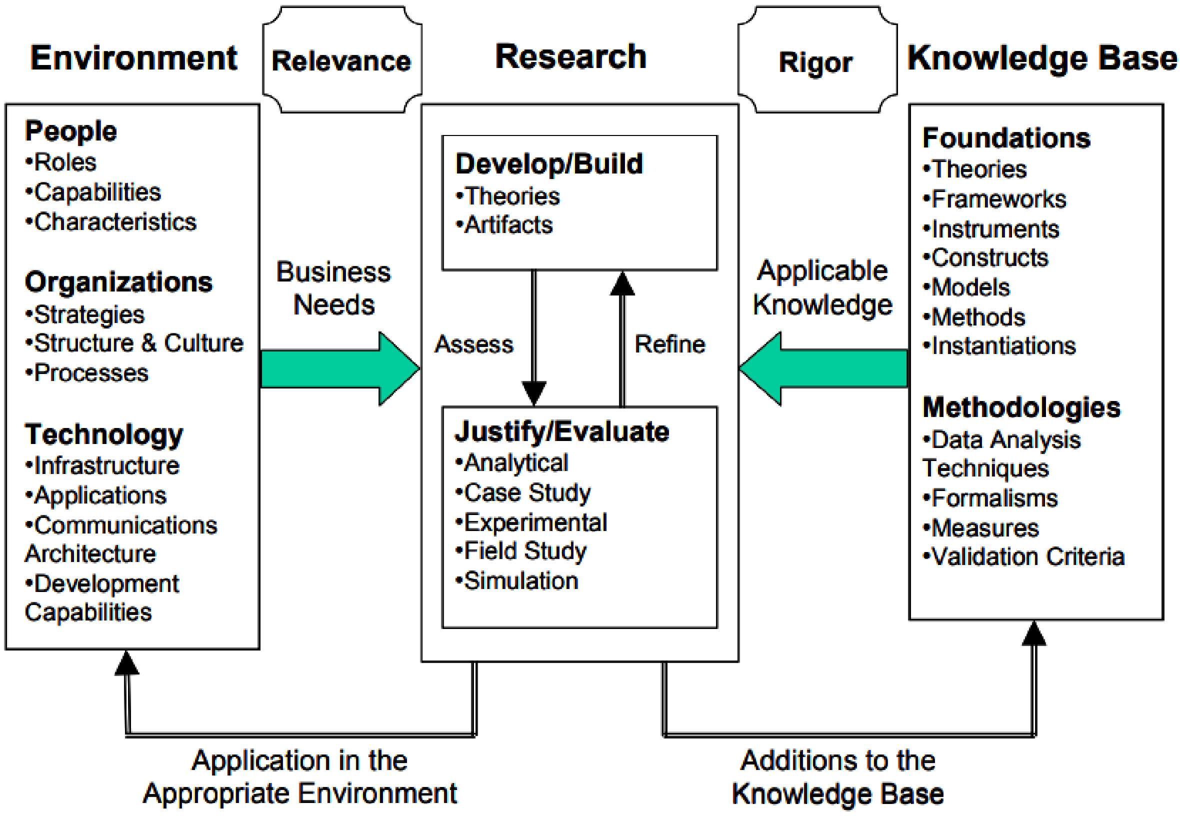


Figure 1.2. Applied Methodology Research Framework (DSR). Source: Hevner, March, Park and Ram (2004)

concerning developing the personality model and personalisation method, meets most of the criteria for research into design experiments (Mettler et al., 2014). These criteria are:

- Controlled testing to assess the quality of the artefact (usability)
- A large number of tested factors and the impossibility of separating them under experimental conditions
- Multiple parallel interactions occurring simultaneously
- Inability to completely eliminate foreign variables - reduction occurs only by simplifying the design (reducing alternative manipulations)
- No standard quality criteria available (should be developed experimentally)

While building a data-based personality model, the researcher operates on difficult experimental conditions such as human nature (personality). The applied methodology treats research as a complex activity based on two essential components: the environment and the knowledge base (see Fig. 1.2). The analysis of the environment is necessary to confirm the validity of the research (meeting the business need and possible application in business practice). This analysis may be based on an analysis of resources, organisation, and technology. The knowledge base analysis is based on a review of existing approaches described in the literature and influences the selection of appropriate research tools. The research must make a clearly identified contribution to the knowledge base.

1.3.2 Steps of the Research

Based on the chosen Design Science framework, the research presented in this dissertation consists of the following steps presented in Fig. 1.3

1. The first step was to identify and define existing problems in electronic service customisation and personalisation market practice. The definition and justification of a specific research problem were done by analysing the gap existing in the service personalisation area. The report from this step is described in Chapters 2 and 3.

2. After the gap definition, a literature review was conducted to search for the possible existing solutions with a detailed analysis of available data. Details of the procedure and process of the literature review are presented in Section 1.3.3.

3. Next, the primary hypothesis about designing the required solution was pre-

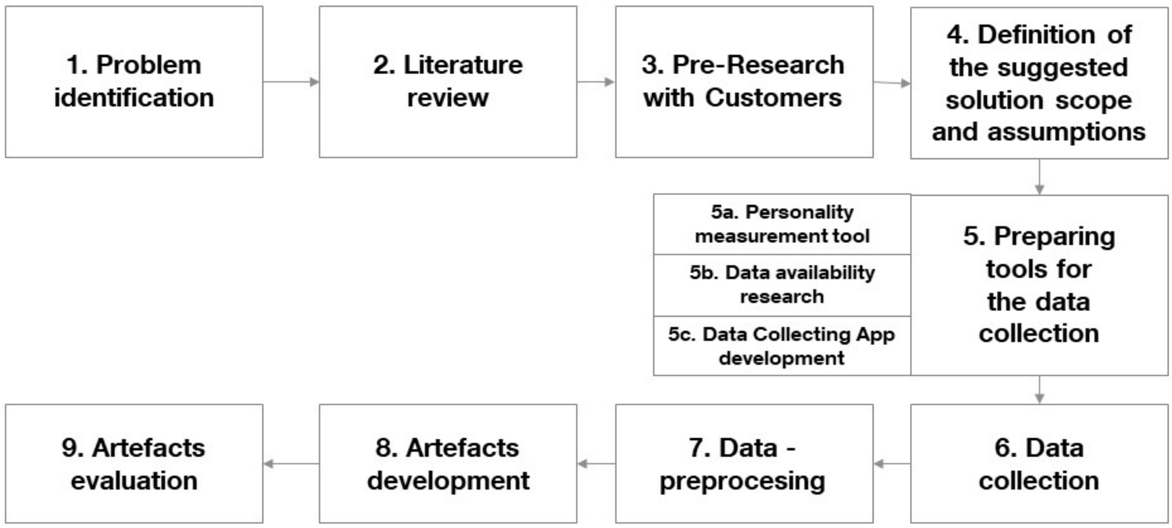


Figure 1.3. Research Steps defined according Design Science Research

(Source: own work)

evaluated in qualitative pre-research conducted on a small sample of Customers - this stage is fully described in Appendix A and synthesised and concluded in Section 3.3.

4. The definition of objectives for creating artefacts is specified based on *state of the art*, knowledge of the current solution, and analysis of business and customer needs. The results of this step are presented in Section 1.2.

5. The design and development of artefacts is preceded by a preparatory stage connected with preparing the environment for conducting development and evaluation step. Because of the procedural requirements at this stage, standard psychometric procedure was conducted to create psychometric tools (Appendix B); it is summarised and discussed in Section 3.4. As the research is based on the data created by customers via mobile phone usage - there was also the need for developing a mobile app for data collecting (Appendix C) and summary of this research part in Section 3.5.1 The data set used for the leading research is presented in Section 3.5.2 with the connection to the list of raw data in Appendix D. The sixth research step is data collecting, described in Section 3.5.2. and Appendix C. The procedure of data cleaning and pre-processing is described in Section 3.5.2, and it is the seventh research step. Then, the appropriate stage for creating artefacts is developed, and this is reported in Chapter 4. The assign-

ment of individual research steps to the stages of the Design Science Cycles is illustrated in Chapter 3.2 and visualised in the diagram 3.1.

6. Finally, the demonstration and evaluation of artefacts was the last ninth step of this research, described in Chapter 5.

7. Communication of the research results and dissemination of the problem, its importance, utility and novelty was executed by publications and participation in domain conferences:

- Data-Based User's Personality in Personalising Smart Services (Krzeminska, 2019)
- Personalizing Smart Services Based on Data-Driven Personality of User (Krzeminska, 2020)
- Personality-based Lexical Differences in Services Adaptation Process (Krzeminska and Rzeznik, 2021)
- Personality Based Data-driven Personalisation as an Integral Part of the Mobile Application (Krzemińska and Szmydt, 2021)

The research has been conducted in three Design Science Research Cycles (1.4. The research started with the Relevance Cycle that defines links between the application domain (Environment) with the research, followed by the Rigor Cycle (between Knowledge Base and Research) and the Design and Evaluate Cycle connected with artefact creation. The Relevance cycle is an iterative process of defining the research problem based on the systematic literature review. At this stage, the researcher formulates requirements for the defined artefact and the criteria for the results' evaluation. This phase was carried out by reviewing existing solutions in personality modelling based on data, service personalisation techniques and survey among customers (Section 3.3). The Rigor Cycle is dedicated to investigating the level of innovation in the designed solution. The main goal of activities in this phase is verification regarding the supplementation of the existing knowledge base in any improvements, extensions or completely new artefacts obtained during the research. For this dissertation, the Cycle involved analysing the existing knowledge base, including scientific theories and methods and existing artefacts found in related fields: management, psychology, and data science.

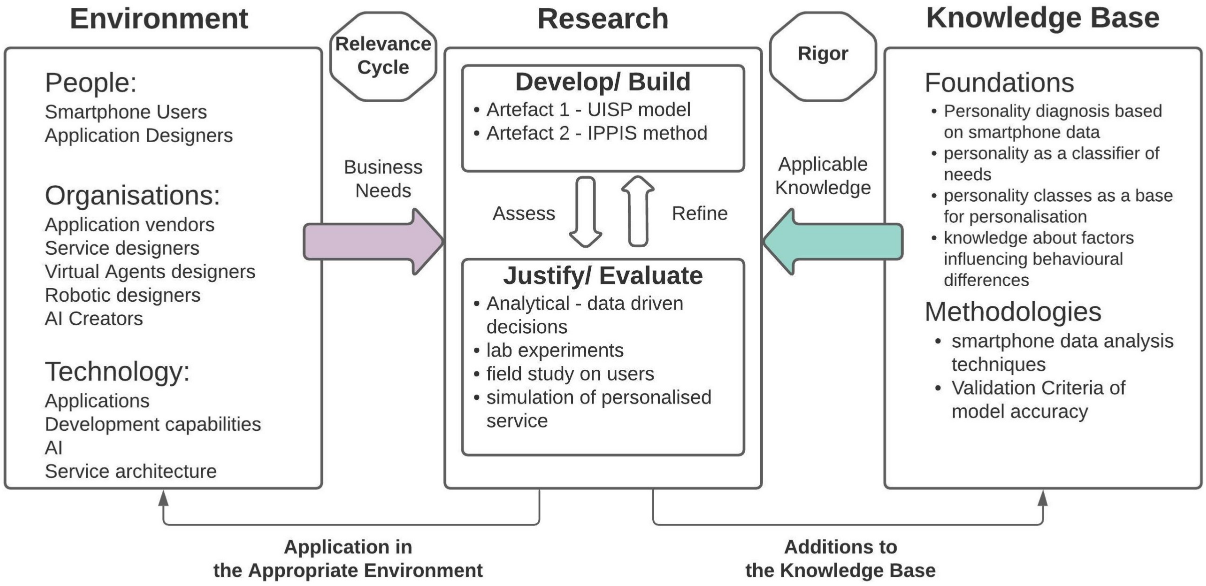


Figure 1.4. Research Cycles defined according Design Science in context of the dissertation.

(Source: own work)

1.3.3 Literature Review Process

Both the Relevance Cycle and the Rigor Cycle required a review of the existing knowledge base. The study covered a wide range of topics because the dissertation concerns the application of psychological knowledge about the behavioural motivations of people in the sphere of personalisation of services, i.e. in the area of business management. Therefore, the literature review covered extensive concepts such as personality theories, personalisation of services and the use of data science and digital traces for personality prediction, as well as the user's right to privacy. The systematic literature review was conducted based on the method proposed by Levy and J. Ellis (2006) and Okoli (2015). In general, this approach consists of three stages of the literature review process: input searching, processing of search results and output creating a list of relevant researches. Input searching is based on keyword search and backward search from references of highly relevant articles.

Then, the literature search was carried out in the following databases: *Google Scholar, SpringerLink, IEEE Database, Research Gate, Academia Edu, Science Direct and Mendeley Search*. A systematic review made it possible to develop a research concept for a defined problem. The keywords were constructed as follows. The primary keys used for the search were:

- *'determining personality based on (digital) data'*
- *'personalisation of services based on the user's personality'*
- *'user's profiling based on data'*
- *'customer-oriented services'*
- *'mobile app personalisation'*
- *'personality as a base of human-robot interaction'*

The search was repeated many times, with the simultaneous replacement of words with semantic synonyms, or British and American spelling:

- 'personalisation' interchangeably with 'profiling', 'personalisation', 'modeling', 'customisation', 'adapting'
- 'user' interchangeable with 'customer'
- 'personality' interchangeably with 'personality profile', 'temperament' and 'Big 5'
- 'data' interchangeable with 'digital data', 'digital traces', 'usage history', 'logs'

The commonness of the words profiling in combination with data, personalisation of services, personality and Big 5 turned out to be a problematic issue returning too many irrelevant articles. Therefore, the limitation to the results (publications, articles) after 2000 was applied. This decision results from the fact that only after 2000 (and even after 2004) digital data began to be used to source information about the user. The most effective was the review of references in publications already classified as relevant.

Table 1.1 shows, for example, the number of articles found only by the Google Scholar search engine for the set of words and the entire phrase. Search results from other search engines gave lower search results and probably matched those from the Google Scholar search engine.

Ultimately, based on the analysis of titles and abstracts, 310 articles related to the searched keywords were qualified for the literature review. Of these, 230 were associated with personalising services, 69 with personalisation, 49 with personality prediction (variously defined), 19 with machine learning methods used for predictive modelling. Additional bibliographic entries were searched for work-related topics such as *management, delay strategy*, and *privacy protection*. Finally, the bibliography for this dissertation contained 574 items. The entire collection (574 bibliographic items) was additionally analysed in terms of the relationship between the topics and the evolution of concepts

Table 1.1. Table of the number of search results found in Google Scholar for sets of words and phrases

searched phrase	a set of words	an exact phrase
determining personality based on data	37200	99
personalisation of services based on the user's personality	22600	0
user's profiling based on data	68600	0
customer-oriented services	11300	3100
mobile app personalisation	8000	1
personality in human-robot interaction	8000	216

(own work)

over time. A graph of the network of connections was created using the VOS Viewer [2] software, thanks to which the dependencies as a function of time are shown.

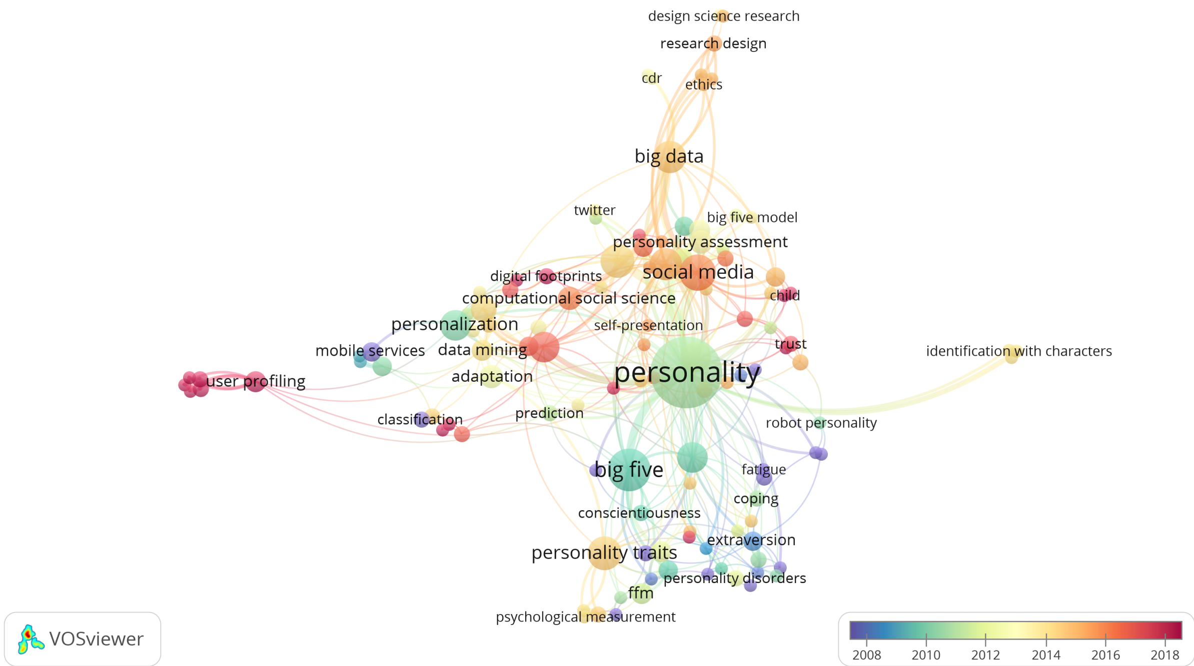


Figure 1.5. Relations between the topics in the bibliography analysed in the dissertation (574 items of the bibliography), the colours reflect the average year for publications for a given topic.

(own work)

On the network graph created basing on the learned literature, it can be seen that personality is the axis of the relevant articles. It is also clear that the concepts related

[2] VOSviewer 1.6.16, http://www.vosviewer.com, Copyright (c) 2009-2020 Nees Jan van Eck and Ludo Waltman

to psychology and personality traits in the psychometric sense are researched were researched even earlier in the 20th century (bottom of the graph, cold colours), and thus the constructs themselves have the lowest average of the year of publication. The dominant yellow-orange area is the cluster already associated with personality detection based on data from social media and computational social science. Yellow indicates that the number of publications in this area of research has been decreasing in the last two years. The graph also shows a separate subject track associated with the personality assessment and social media cluster for big data. Big data, in turn, is related to the issue of ethics and the construction of the research itself. The graph also shows that the latest areas related to personality determination (an average year of publication is 2018) are related to user profiling, data mining, learning and classification. It turns out that the graph of connections between topics, combined with publication dates, illustrated the direction in which the entire area of research is heading - i.e. personalisation using advanced prediction techniques.

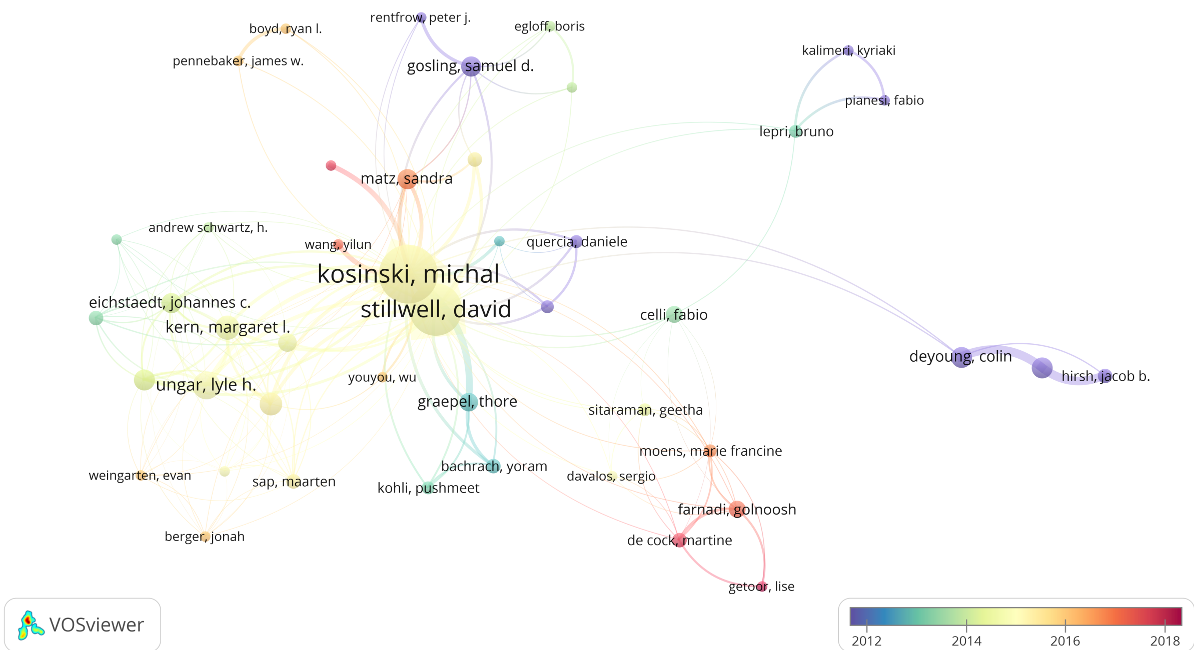


Figure 1.6. Relations between authors in the bibliography analysed in the dissertation (574 items of the identified bibliography), the colours reflect the average year for publications from a given author. Only authors with the highest frequency rates are presented on the graph

(own work)

A similar analysis carried out on the connections between 1242 authors of research and publications shows how much influence Michał Kosiński and David Stillwell had on building the knowledge base in identifying a personality based on data (Figure 1.6). Probably, if it were not for the MyPersonality App, which collected data for this research and enabled very large-scale analysis, this area would not have gained such popularity. Moreover, the researchers who started work on the MyPersonality App project, collecting digital data from Facebook and Twitter, are now continuing their research beyond social media. It turns out that the graph of connections between topics, combined with publication dates, illustrated the direction in which the entire area of research is headed - i.e. computational personality (Golnoosh Farnadi, Marie Moens, Martine de Cock) and personalisation using machine learning.

There are under 50 articles that meet all the criteria, and most of them relate to social media use before 2014 and the Cambridge Analytica [3], connected with App gathering Facebook data, was commonly evaluated as unethical ((Isaak and Hanna, 2018),(Laterza, 2018)) and significantly change the availability of social media data and the way digital footprint are treated and considered. On the one hand, unethical cases of using data resulted in greater control and legal procedures based on consent. On the other hand, the encapsulation of access to data was limited mainly to one service. However, no legal legislation will prevent the useless use of data, which has become one of the motivations for the research presented in this dissertation.

Finally, it should be noted that data-based personalisation of services is an area where science meets business, which is not conducive to knowledge sharing. Some solutions can be created and tested and will not be made public until they are granted patent protection. This work is based on a review of the knowledge base available in publications.

1.3.4 Defined Artefacts

The third Design Science cycle is Design Cycle, which resulted in designing and evaluating artefacts. Taking the Design Science criteria into account and the documented

[3] cite from Wikipedia *'The Facebook–Cambridge Analytica data scandal concerned obtaining the personal data of millions of Facebook users without their consent by British consulting firm Cambridge Analytica, predominantly to be used for political advertising. The data was collected through an app called "This Is Your Digital Life"*, developed by Global Science Research in 2013. The app consisted of a series of questions to build psychological profiles on users, and collected the personal data of the users' Facebook friends via Facebook's Open Graph platform.' from https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

experience of other researchers work, as well as the business motivation presented in Section 1, the research program was created to accomplish the project of creating user-oriented smart services. This program aims to evaluate two defined artefacts:

- **Artefact 1:** A model of determining a user's initial personality profile based on the minimum amount of data available at the time of service installation
- **Artefact 2:** A method of implementing a personality profile for automatic personalisation of the service (based on example of mobile app)

It should be noted that the first artefact was called a model due to its structure and origin. Ultimately, it will be a kind of statistical model based on probabilities, created thanks to automatic machine learning techniques. In turn, this model (Artefact 1) can be used by the automatic method of building the personalization logic (Artefact 2). And this method can be adapted to various electronic services, in which there is an interaction with the user.

Regarding the research contribution to the knowledge base, both of the defined artefacts can be classified at least as research opportunities (A. Hevner and Chatterjee, 2010). Although, through the literature review, no solutions similar to the second artifact were found, which may indicate low application domain maturity and innovation of the solution. On the other hand, the application domain is closely related to the business sphere, which means that similar solutions may exist but they are protected against possible competition, and not announced in scientific publication. The demonstration and evaluation are pivotal steps in the last Develop and Evaluate Cycle. The designed artefact should be assessed by methods that fully demonstrate its quality, effectiveness and usefulness (Hevner 2010). Chapter 5 describes the artefact evaluation process in detail. The artefacts were assessed using the Framework for Evaluation in Design Science Research (FEDS) proposed by Venable et al. (2016). According to the classification proposed by FEDS, the evaluation applied to the research presented in the dissertation is an ex-post evaluation, in line with the risk and effectiveness strategy envisaged for the created technical implementations whose recipient is the user. Moreover, for the current artefacts, both an artificial (criteria analysis, simulation) and naturalistic method of assessment (test in action and customer evaluation) are provided. A separate artefact evaluation process and evaluation criteria have been developed for each artefact (Chapter 5).

1.4 Structure of the Dissertation

In the context of this dissertation, an environmental analysis was conducted: review of personality theoretical approaches (Section 2.2), implications for dissertation purposes (2.2.2) and the relation between personality profile and classification of human needs, in electronic services' context, based on results of the preliminary research (Section 2.2.3). As for the knowledge base, a brief introduction to the psychological theories of personality (Section 2.1.2) and the origin of the theory used in dissertation are compiled (Section 2.1.3). And due to the dissertation thesis the related extensive literature review for the techniques and methods of determining personality based on data is presented in Section 2.2.1. The entire research was justified by the research needs presented in the motivation (Section 1.1) and the identified shortcomings from literature review (section 2.6) and preliminary research (Section 3.3 and 3.5). Concept of Personality Aware Service in which both artefacts can be used is presented in Section 2.8. To respond to these needs, relevant artefacts were created developed, by used of the tools and techniques described in sections 2.3 and 3.4. Then, process of artefacts creation is described, step by step, in Chapter 4. Finally, the artefacts evaluation according to the FEDS is presented in Chapter 5.

Chapter 2

Personality and Role of Personalisation in Interactive Services

2.1 Introduction

This chapter synthetically presents the knowledge base that is the foundation for designing the entire research process. It consists of two main parts, one dedicated to personality (theories and possibilities of indicating it from digital data) and one dedicated to the penalisation methods, roles, and limitations. Since personality is the key and distinctive element of the method, it is presented in Sections 2.2.2 and 2.2.3 after a synthetic discussion of the theory of personality (Big 5) against the background of personality research history, other approaches, and their assumptions (Section 2.2.1). In the next section, the Big 5 personality theory is discussed in terms of the premises for treating it as a reflection of human cognitive needs (Section 2.4). Then, examples of existing research and publications on the determination of user personality directly from the data (2.3.1) will be discussed. Subsequently, the preliminary research results of the suitability of personality to classify the user's needs of an electronic service is presented (Section 2.3.2). Personality role in human-service interaction is discussed in Section 2.4, which presents studies demonstrating the relationship between the user's personality and how the user interacts with interactive services based on artificial intelligence (mainly on the example of interaction with robots). Before presenting a new idea for personalising services (2.8), the definitions and types of personalising and customising services are discussed, and the advantages and disadvantages of currently used methods of personalising interactive electronic services (Section 2.5).

2.2 Personality Theories Review

Personality does not have a single definition commonly shared by all psychologists. The history of personality theory is older than the history of psychology as a science. The first founders of a kind of personality typology that can be mentioned are ancient philosophers like Hippocrates, Plato, and Aristotle, since all of them describe different types of humans. It was further explored by other great philosophers and by scholars who tried to define concepts related to humans. Psychology developed at the end of the 19th century as a derivative of philosophy (human concepts) and experimental physiology. True personality theories, based on a pure psychology basis, began to take shape at the beginning of the 20th century due to clinical observations of doctors - psychotherapists, practitioners, not scientists (Freud, Jung, McDougall). Nowadays, personality is defined as a set of traits or features that are organised, relatively stable in time, influencing the cognition and behaviour of individuals (Strelau, 2019; Strelau and Zawadzki, 2012). The relationship between theory and practice determines the distinctiveness of personality theory against other psychological theories. Personality theories have always been functionally oriented, related to personality measurement, and focused on problems relevant to organism's adaptation and individual survival. Personality theorists attached the most importance to motivational processes and treated them as the key to human behaviour. This already shows the primary personality feature that is crucial for this work, i.e., its vital role in determining behaviour. The table shows the main concepts and theories from a historical perspective (see Tab. 1). The list is incomplete but it's an attempt to show various directions dependent on therapeutic approaches. Allport (1937) distinguished 50 different definitions of personality in a literature review. It is also worth noting that many personality researchers (psychologists) try to answer two primary research questions within the adopted assumptions concerning the concept of human functioning:

- What makes people behave similarly?
- What makes people act differently?

Researchers are looking for the minimum number of conditions, factors, and variables to explain the same responses in all people. At the same time, they look for the reasons for the differences in imbalances in response to the same situation. Every personality researcher is looking for individual differences.

Table 2.1. Comparison of different personality approaches with basic assumption concerning structure (prepared based on the book 'Theories of Personality' by Hall C.S. (2004))

Approach	Creators	Techniques	Personality Structure
Psychoanalysis Theory	Z. Freud (since 1900)	Case studies, self-analysis, free associations, dream analysis	Id, Ego, Superego
Analytical Theory	C.Jung (since 1921)	experimental research on complexes, case studies, comparative research (searching for inter-culturally shared archetypes), dreams	ego, personal unconscious, collective unconscious, self, attitudes, functions, relationships between personality elements
Psycho-social theories	A.Adler, E.Fromm, K.Horney, H.S.Sullivan, E.Erikson (since 1938)	childhood experience study, birth order impact study, interview, personality disorder studies, social interests	dynamism, personifications, cognitive processes, purpose of life, interpersonal relations, personality as illusion (hypothetical being)
Personology	H. Murray (since 1938)	in-depth research in groups of undisturbed people diagnostics, the first tools for measuring personality, research on representative samples	personality definition, courses and series, time distribution, abilities and achievements, personality systems
Idiographic approach	G. Allport (since 1954)	idiographic and nomothetic approach, direct and indirect personality measurement, studies on expression	personality, character, temperament, trait, intentions, functional autonomy, unity of personality
Factor theory of traits	R. Cattel (since 1945)	studying individuals with factor analysis, model of personality, specification equation	the structure of traits, traits, ability and temperamental traits, dynamic features
Biological Traits Theory	H. Eysenck (since 1947)	psychological models explaining the causes, factor analysis, experimental psychology connected with neuroticism and extraversion	Temperament description; extraversion, neuroticism, psychoticism

(own work)

The Tab. 1, which is based on the review of the book of Hall C.S. (2004) contains most of the main theories of personality, focused on theoretical constructs concentrated on the person and linking personality with motivational mechanisms and thus influencing behaviour. Ideas omitted include i.a. theories related to the general idea of character psychology, which relate to cognitive functions such as perception of reality, learning, and thinking (George Kelley's theory of personal constructs or theories highlighting the learning phenomenon, comparable to other researchers: B. F. Skinner, Dollard, Miller or Albert Bandura).

It should be noted that all theories of personality emphasise the elements that differentiate individuals. However, due to the defined research goals, theories that do not concern behaviour can be ignored at the outset. For example, psychodynamic theories (Freud, Jung, Horney, Fromm, Sullivan Erikson) focus on the unconscious: motives of behaviour and interpersonal conflicts. In turn, structural theories (Murray, Allport, Cattell, Eysenck) mainly concern behavioural tendencies that characterise individuals.

Based on Freud's psychoanalytical theory, the Swiss psychoanalyst and psychiatrist Carl G. Jung (1875-1961), the creator of Analytical Theory, formulated a theory about two types of mental attitudes: extroverted and introverted. Extraversion means a tendency to direct energy and attention towards objects of the external world, and introversion means directing it towards the inner world (internal experiences, ideas, concepts, ideas). Two American psychoanalysts from the school of psychoanalysis, Katharine Cook Briggs (1875-1968) and Isabel Briggs Myers (1897-1980), created a typology of personality, the so-called MBTI (the Myers-Briggs Type Indicator) (Isabel Myers, 2010). It is an array of 16 personality types: INTJ, INTP, ENTJ, ENTP, INFJ, INFP, ENFJ, ENFP, ISTJ, ISFJ, ESTJ, ESFJ, ISTP, ISFP, ESTP, and ESFP), which are the result of a combination of 4 binary dimensions:

- Introvert (I) or Extrovert (E) shows how an individual interact with each other;
- Intuition (N) or Sensing (S) shows how an individual see the world and process information.
- Thinking (T) or Feeling (F) shows how an individual make decisions and cope with their emotions
- Judging (J) or Perceiving (P) reflects an individual's approach to work, making decision, and planning

The MBTI approach is now more of a product commonly used in human resource management and is often used to identify strengths and weaknesses in the context of work psychology.

Although related to the aim of the study, the psychodynamic approach concerns empirically unverifiable, unconscious determinants of behaviour. They are unverifiable because, as a rule, they are immeasurable (unawareness). It was only with the works and research of Murray, Cattell, and Eysenck that the evolution of the theory of human personality and attempts to link individual differences with genetic and biological factors began. An increasing importance was also attached to the measurability of personality indicators (distribution of characteristics in the population). The work and research of Murray and Cattell began with a rich history of standardised questionnaire tools for measuring personality. For the needs of these theories, reliable statistical methods were constructed, e.g., factor analysis. Henry Murray was the first to try to define personality: *'the history of personality is a personality'* (Murray and McAdams, 2010). Murray understands personality as a factor that organises and guides a person. Thus, personality is directly related to the brain's work (Hall C.S., 2004).

It is also worth noting that personality theorists, particularly those for whom personality is trait-based, closely associate it with the biological teachings and predispositions of the human nervous system. It is worth emphasising the entirety of Catell's research and theory, from which the Big Five grew, and the importance of Eysenck's research, which was the first to point to the biological basis of 3 features of the big five (Extraversion, Neuroticism, Agreeableness).

The anthology 'Theories of personality' (Hall C.S., 2004), it is even said that personality theorists as a group are more biological than social. Personality research in the context of human behaviour motives (Cattell, Murray) attempts to translate the differentiation of behaviours into a set of specific variables explaining this differentiation. One can risk a statement that they are the best theories due to the highest numbers of research endeavors and experiments and, in this way, becomes a well-documented measurability and a connection with statistical descriptions of characteristics in the population.

The presented characteristics of the personality theory approach based on traits against the background of the existing theories of personality allow for the identification of its most important attributes. First of all, the fact that theories of features are related to the biological nature of humans is strongly supported in research on the entire population and not only on individual cases (psycho-dynamic theories).

Within the framework of currently existing traits' theories of personality, one of the most influential and quantifiable is the Big 5 model. The model was derived from Cattell and Eysenck's works (compare Eysenck (1981, 2019), mainly developed by Costa and McCrae since 1978 (Costa and McCrae, 1980). In the 1990s, it was confirmed in many empirical studies (Costa and McCrae, 1992, 1995; De Raad, 2012). Many studies indicated a strong relationship between personality and behaviour, life satisfaction, as well as achievements and preferences (Barrick and Mount, 1991; Judge and Ilies, 2002). Detailed characteristics of this theory and the basis for selecting it as fit for this study will be presented in the next chapter.

2.2.1 Big Five Origin and Assumption

The Five-Factor Model of Personality, also popularly known as the Big Five, is the dominant approach to the basic dimensions of personality in psychological research. It has also been the most discussed approach in recent decades. The interest in personality taxonomy began in the 1960s. Several personality scales were then analysed, including those distinguished by Cattell (Extraversion, Stability). Norman was the first to show in 1963 that personality traits are grouped into five factors (Norman, 1963). The first research was mainly based on the analysis of adjective lists. In his work, Norman (1963) (p. 577) gave the factors the following names:

- factor I - Extraversion or Surgency
- factor II - Agreeableness
- factor III - Conscientiousness
- factor IV - Emotional Stability
- factor V - Culture

In the literature, the factors are often differently named in many researches. However, their quantity does not change - there are always five. That is why the concept was called the 'Big Five.' The theoretical model was further developed based on two research trends: lexical and questionnaire (De Raad, 2012; John and Srivastava, 1999;

Ruch et al., 1991). Although the factors themselves differ from one another in the history of various studies based on lexical and factor analysis, there is also a great agreement among the five-factor models themselves. According to this approach, there are five basic dimensions of personality. They are: Extraversion, Emotional Stability (also called Neuroticism), Intellect or Openness to Experience, Agreeableness and Conscientiousness (Costa and McCrae, 1992; De Raad, 1999; Goldberg, 1999; John and Srivastava, 1999; Trapnell and Wiggins, 1990). The research focusing on searching for fixed and measurable personality dimensions was based on the lexical hypothesis, according to which all essential aspects of personality were encoded in languages. The main goal of these studies, which started in the 1950s (Digman, 1990; John and Srivastava, 1999; McCrae and Costa, 1987, 1997; McCrae and John, 1992) was to isolate the mappings in language elements (mainly adjectives) and, based on these mappings, create a structure reflecting human personality. This ordering took the form of a gradual reduction of a large set of words describing the human personality to a smaller set of dimensions. The most frequently used analytical method in these studies was factor analysis. Subsequently, the studies were developed and continued in the questionnaire trend, where the generated five-factor theoretical model was shaped conceptually and empirically. The five dimensions of personality were operationalised in questionnaires, the relationships of these dimensions with other models and theoretical constructs were investigated, and the implications of this model for the key issues in personality psychology were considered and discussed. In this line of research, five factors were no longer treated solely as defining features. They began to be interpreted as tendencies or dispositions to specific reactions, feelings, and behaviours. Furthermore, a theoretical description for the five distinguished factors was created (Tab. 2).

Since the 1990s, the Big Five model has been confirmed and psychometrically verified in many empirical studies (Digman, 1990; John and Srivastava, 1999; McCrae and Costa, 1997, 2004; McCrae and John, 1992). The results of many of these studies have confirmed the legitimacy of distinguishing the Big Five taxonomy. The Big Five is the most commonly used model to describe human personality. There are also several studies looking for the relationship between the Big Five Factors and temperament. Researchers disagree and consider them as factors of both personality and temperament.

It all depends on the definition and assumptions. According to Angleitner (1991, p.

Table 2.2. Descriptions of the Big Five factors and the main activity to which they relate, based on (Strelau, 2019)

Dimension	Description
Factor I - Extraversion	Determines tolerance to external stimuli (quantity and intensity). It manifests itself in a tendency to be open, assertive, active, and looking for emotions
Factor II - Agreeableness	Defines tolerance for other people - a tendency to be polite, gentle, trusting, and to give up one's own opinion/goal
Factor III - Conscientiousness	Specifies a tendency to focus on work and goals - a tendency to be thorough, responsible, organised, diligent, achievement-oriented, and persistent
Factor IV - Neuroticism or Emotional Stability	Determines tolerance to stress - a tendency to anxiety, fear, depression, and mood, overreacting
Factor V - Openness to experience or Intellect	Specifies tolerance to new experiences / unknown situations a tendency to creativity, imagination, experimentation, perceptiveness

(own work)

190), *'I see the first four factors of the Big Five as primarily temperamental dimensions.'* In this approach, temperament is defined as similarly to the regulatory theory of temperament by Strelau (1998, p.34) *'Despite differences in conceptualising this phenomenon [temperament], most temperament researchers agree with the following definition: temperament refers to basic, relatively stable personality traits that are present from early childhood, occur in human beings, but also have their counterpart in animals. Being primarily determined by inborn neuro-biochemical mechanism, temperament is subject to slow changes caused by maturation and individual-specific, genotype-environment interactions.'*.

On the other hand, Hofstee (1991, p.184) for whom temperament is the biological (gens, glands, and nerves) core of personality claimed that *'The more temperamental a trait is, the more fundamental it is judged to be with respect to the concept of personality.'* In Hofstee's research, the highest levels of fundamentality, understood as naturalness (coming from nature), were achieved by factor I, i.e., Introversion (equivalent to extraversion), and factor IV, i.e., Emotional Stability (Neuroticism in other models). This *naturalness* can also be seen in the statistical distribution of features (raw data distribution, Appendix B). The cumulative distribution function for Extraversion and Neuroticism before normalisation is similar in shape to that of a normal distribution. The remaining features have strongly skewed distributions and need to be normalised.

In the same article, Hofstee argues with Strelau's line of reasoning that personality is at least in part socially determined. Hofstee (1991, p.183) claims that *temperament does not coincide with personality, but is a proper subset of it ... temperamental traits constitute a significant subset rather than a peripheral one*. Also, he proposes to define the following differentiation between temperament and personality:*'Temperament is biologically determined where personality is a product of the social environment'*.

This classification of the Big Five factors as temperamental dimensions is not without significance for their subsequent use. Considering the purpose of using this classification, the biological basis of at least some of the Big Five dimensions would be the confirmation of their biological conditioning, and thus their relative stability over time. The widespread introduction and use of the Big Five measures based on factual data will significantly affect the availability of data necessary to validate these hypotheses.

The theoretical model of the Big Five has repeatedly confirmed its usefulness in predicting functioning in various areas of life, ranging from psychological diagnostics, career counseling, and activities in the area of human resources by targeting communication at specific groups of recipients. Models, concepts, and systems based on this research have gained their separate commercial names, such as Canoe and Ocean (anagrams from the first letter of traits names).

All this points to the comprehensive and universal nature of the five-factor model and its high usability. On the other hand, the biological foundations and theirs relation to the context of the human nervous system reactivity led to the comparison of this model with the universal physics of personality (Allik and McCrae, 2006; Allik and Mccrae, 2004; McCrae and Costa, 2012). In this trend, research is still being conducted, with the overarching goal of creating an integrating and universal model of human functioning based on personality or temperamental traits, emotions, motivations, and social characteristics (Strus and Cieciuch, 2014).

The fact that this area has active continual research means that the model is not complete, and the scientific consensus on many issues has not been reached (e.g., at the questionnaire level, the problem is the lack of orthogonality of factors, which may result from many issues, and one of them is, for example, declarative questionnaire descriptions). The lack of sufficient empirical data does not even allow the Big Five factors to be unequivocally classified as a temperament or personality model. The development of technology and easier access to data, recording and reflecting actual behaviour, and not declarations about oneself, as is the case with questionnaire surveys, have opened up some new research fields.

2.2.2 Big Five Description and Implications

The previous Section presents the terminological and theoretical controversies connected with the Big 5 theory. For these studies, four criteria were adopted that had to be met in order to consider a factor as a personality dimension (after Costa and McCrae (1992)):

- The reality of the factors manifested at constancy, accuracy measured by the compliance of observers' opinions and practical utility
- Diffusion or the fact of appearing in multiple forms in all concepts of personality
- Universality that is, occurring regardless of gender, age, race, and culture
- Biological substrate

A large body of scientific evidence has confirmed that the Five-Factor model fulfills those criteria. The biological basis of the Big Five dimensions has been mentioned earlier in this chapter (see also: Angleitner (1991), Goldberg (1992), Strelau (1998), Strelau and Zawadzki (2012), and Zawadzki (2017)). The Big Five dimensions appear cross-cultural independent (criterion of universality), as the same five factors have been identified in many studies all over the world (Carlo et al., 2013; Triandis and Suh, 2002). The dimensions of the Big Five also accurately predict behaviour (criterion of accuracy). For example, coexistence in a person's profile of: high Conscientiousness, low Neuroticism, and high Agreeableness is good predictor of successful job performance (Ones et al., 1994; R. P. Tett et al., 2000; R. R. Tett et al., 1994).

Big Five dimensions also predict the performance of leaders; ratings of Openness to Experience are positively correlated with ratings of leader success, while ratings of Agreeableness are negatively correlated with success (McCann, 2005; Rubenzer et al., 2000). The Big Five factors are also increasingly used to help researchers understand the dimensions of mental disorders such as anxiety and depression (Saulsman and Page, 2004; Zawadzki, 2017). As mentioned earlier, majority Big Five researchers claim that this trait is biologically determined.

The advantage of the five-factor approach is that it is economical and straightforward. Instead of studying hundreds of features, researchers can only focus on the five fundamental dimensions. The Big Five can also capture other dimensions that have been of interest to psychologists. For example, the need for achievement relates to the dimension of Conscientiousness, and self-esteem relates to low Neuroticism. While 5 great factors have a significant influence on behaviour, they do not consider all essential dimensions of personality. For example, the Big Five does not capture moral behaviour, although this variable is vital in many personality theories. In addition, there is evidence that the factors of the Big Five are not precisely the same in all cultures (Cheung et al., 2001; McCrae, 2002).

The Big Five model results from decades of work using factor analysis to map covariance patterns between features. More precisely, such factor analysis require a relatively comprehensive and impartial feature pool to be analysed. The lexical hypothesis states that natural language (presented in dictionaries) provides just such a pool of feature descriptors (John et al., 2008; John and Srivastava, 1999; Saucier, 2009). Existing personality questionnaires are another source of a large and broad pool of traits in which behavioural descriptors can be located. Lexical and questionnaire studies provided evidence for the Big Five (Digman, 1990; John et al., 2008; Markon et al., 2005). However, explanatory models for the Big Five should be possible because the fundamental premise of the factor-analytical approach is that many features differ, pointing to a factor, due to a common cause (Haig, 2005). The Big Five theory seems to have a dominant position in personality research, and for many important reasons, it is also used for creating new Cybernetic Big Five Theory. In this approach, a personality trait is defined by DeYoung (2015) as *'probabilistic description of relatively stable patterns of emotion, motivation, cognition, and behavior, in response to classes of stimuli that have been present in human cultures over evolutionary time'*.

The proposed Cybernetic Big Five is organised in the same way as the traditional Big Five model and consist of five-dimensional factors: **Extraversion** (tolerance for big quantity of stimuli and high need for socialising), **Openness to experience or Intellect** (tolerance for the new and the unknown), **Neuroticism** (intolerance for stress), **Conscientiousness** (intolerance for chaos and disorder, **Agreeableness**(concentration on others' needs and willingness for co-operation) (DeYoung, 2015). The above literature review shows that the Big Five model is a valuable model that has been proven in many experiments to differentiate human behaviour. Moreover, it seems that it can also classify people according to their needs and motivations.

2.2.3 Big Five as a Classifier of Human Cognitive Needs

As mentioned before, the Big Five personality model is widely used and confirmed by many empirical studies (Costa and McCrae, 1992; De Raad, 2012; Goldberg, 1992; McCrae and Costa, 1996). In addition, several studies show a strong relationship between personality and behaviour, life satisfaction, and achievements and preferences (Barrick and Mount, 1991; Judge et al., 1999).

Matz et al. (2017) confirmed that the use of psychological targeting allows influencing the behaviour of social media users by adapting the message to the psychological needs of recipients defined by personality (Big Five). However, at the beginning of using a new type of AI service, such as a virtual/voice assistant, data from the history of usage will not be available. In the context of even better regulations on protecting users' privacy, it can be observed slow change towards a situation in which services will be limited to the use of their data, and data information systems will become more and more hermetic. Therefore, more often there is a need for methods that will allow services to detect the user's personality based on even small amounts of data available at the time of service installation, e.g., from a mobile phone (apps used, history of usage, notes, calendar, photos, way of data storage).

In this paradigm, the assumption refers to determining the personality of each user of a smart device (personal or home), so the model should be based on the data available from different smart devices. Therefore, such data types that do not apply to everyone (e.g., social media or other specific services or applications) was excluded .

Each subsequent technology and subsequent intelligent services will be increasingly complex and challenging to learn. It seems that the adaptation process will be more and more difficult. Understanding a user's personality at the beginning of the adaptation process seems crucial for applying an appropriate activity to eliminate cognitive barriers.

The adaptation process is more straightforward for those with high Extraversion, high Conscientiousness, and low Neuroticism (Matthews, 2008). Highly neurotic people are generally not resistant to stress, accompanied by a higher level of anxiety and a lower ability to adapt, which stands in contrast to high Openness, as it is a beneficial feature when learning about and discovering new things. An additional incentive for people with high Openness to using such technologies is their intellectual involvement. According to Matthews, Hancock, et al. (2020) and Matthews, Lin, et al. (2020) robots and virtual agents are simulating human behaviour better and better to facilitate contact and cooperation (rich communication, support, kindness, warm and deep contact). Simulation of human nature promotes the anthropomorphisation of virtualised entities.

Interestingly, people with high Extraversion and Emotional Stability (low Neuroticism) anthropomorphise robots and virtual agents more often. They create emotional closeness, which facilitates contact and cooperation (Salem et al., 2015). This dependency also works the other way round; robots with highly extroverted behaviours give a more favourable impression than robots with introverted features (Robert, 2018), which can be a good solution in the absence of information about the user's personality profile. In turn, research (Q. Zhang et al., 2019) shows that contact with robots or other intelligence simulations can be facilitated by mapping a user's personality. Extroverts prefer robots with extroverted behaviours and introverts with introverted ones.

In conclusion, in the light of the presented experimental research in the field of psychology (subsection 2.1.2 and 2.1.3) and given its constant usefulness in explaining individual differences in behaviour (subsection 2.13 and 2.1.4), the Big 5 personality model seems to be a good criterion for determining users' needs. Moreover, it seems that it can be used for research to create automatic personalisation based on digital data from users.

2.3 Methods of Personalising Services

As mentioned at the beginning of this dissertation, one of the main motivations was to use the user's available digital data for automatic classification of needs and automatic adaptation of digital and interactive services to the user. After introducing personality psychology, the primary purpose of this chapter is to review the existing solutions for the automatic determination of the user's personality according to the Big Five models from the available digital data. The methods of determining the user's personality based on data are implemented to obtain in-depth, valuable insight about the user or the client of the service. This knowledge is to help matching services and communication (advertising) better to the client's needs. An example of such an application of user-tailored advertising is presented in the article (Matz et al., 2017). This article analyses the effectiveness of such personalisation with a positive result. First, a research gap will be described in the context of determining personality from data. Moreover, in the next step, the gap in terms of the existing methods of customising services to the user's needs will be identified.

2.3.1 Existing Methods of Indicating Personality Based on Data

Over the past ten years, many researchers and companies have tried to determine a user's personality from different types of digital data. Most of the trials involved data from speech (Polzehl, 2015), social media like Facebook, Twitter (Kern, Eichstaedt, Schwartz, Dziurzynski, et al., 2014; Liu et al., 2016; Quercia et al., 2011; Quercia et al., 2012) or other personal data such as call logs (Montjoye et al., 2013), mobile applications (Frey et al., 2017; Xu et al., 2016) or digital heterogeneous data (Wei et al., 2017). There is also evidence that this type of data-based personality identification is accurate (Back et al., 2010; Kalimeri et al., 2013). Regarding the methods used, the data-driven user personality models are mainly based on text analysis (e.g., Tweets or FB posts) using statistical regression or simply Pearson-r correlation (Golbeck et al., 2011; Kern, Eichstaedt, Schwartz, Park, et al., 2014), and partially supervised machine learning (Back et al., 2010).

Montjoye et al. (2013) proved that it is possible to predict personality from data from telephone logs. There is also a proven case of detecting personality based on socio-metric tokens after collecting various data types reflecting human behaviour for six months (Kalimeri et al., 2013). Finally, one of the most recent studies Berkovsky et al. (2019) concerned the detection of personality traits from eye-tracking[1] data as an alternative to paper-based personality assessment tools. Notably, this research aimed to develop a clinical personality tool that requires much better accuracy than marketing tools.

[1]Eye tracking is measuring the point of gaze and tracking the movement of the eyeball in relation to the head. This measurement is usually carried out by a special device placed on the head of the examined person or by an external system based on sensitive cameras analysing the movement.

Digital footprint tracking to detect users' personalities has been extensively studied by researchers using all kinds of large datasets such as text, profile picture, music, movie preferences based on FB likes or stories (SNA), for example, (Kern, Eichstaedt, Schwartz, Dziurzynski, et al., 2014; Kern, Eichstaedt, Schwartz, Park, et al., 2014; Liu et al., 2016). However, aside from social media profile photo analysis, all other studies were based on massive data gathered from social media usage history. Nevertheless, at the beginning of using new kinds of AI services like virtual / voice assistant, usage history data will not be available.

In the context of even better and stricter regulations on protecting user privacy, services will only be limited to using their data collected during the use of the service, and user information systems will become increasingly airtight. Therefore, there is a need for creating methods that will allow services to detect a user's personality based on even a small amount of data available at the time of service installation, e.g., from a mobile phone. In this paradigm, the assumption is about determining each user's personality of the intelligent home service. Consequently the model should be based on the data available on each phone. As a result, data types that do not apply to everyone (e.g. social media or other specific services or applications) was excluded. The tables below (Tab.2.3, 2.4, 2.5, 2.6 2.7, 2.8, 2.9, 2.10, 2.11) presents relevant research articles describing the techniques used to detect personality based on the different types of digital traces. The tables also conclude the main advantage and limitation of the research.

Table 2.3. Comparison of different methods of determining personality profile based on various data - part 1

No	Paper name	Techniques	Advantages	Limitations
1	Predicting Personality from Twitter (Golbeck et al., 2011)	Big 5 model derived from social media texting, N=50, usage of psycho-linguistic tools	the best performance achieved for Openness and Agreeableness, error 10 p.p to real value, validated by FB/Twitter add campaigns (Matz et al., 2017)	requires 2000 tweets per user to achieve; some language features were not considered during analyses e.g. misspelled words in tweets; only r-Pearson is published in the article
2	Our Twitter Profiles, Our Selves; Predicting Personality with Twitter. (Quercia et al., 2011)	Big 5 derived from social media publicly available activity r- Pearson correlation, web app, users give their consent to share their personality score and profile information	the usage of three counts (following, followers and listed counts) predict user's personality better; validated in (Matz et al., 2017)	the issue of people creating fake accounts or information; the prediction traits are made informally based on intuitions and thus they can'not guarantee a good level of accuracy.
3	Mining user personality in Twitter (Celli, 2011)	Big 5 from individually written Tweets, USML, based on 12 extracted linguistic features, calculated for anyone with more than 1 tweet.	validation between different social-media performance of the same user	personality treated as a binary variable, accuracy is calculated for coherency of personality between various social media.
4	The role of emotional stability in Twitter conversations (Celli and Rossi, 2012)	linguistic differences on personality, USML: statistic-based	investigation of the difference between the stable and neurotics, neurotics posts more and have longer exchange chains	analyses are focused on re-tweet ratios, this is no exact personality prediction
5	Unsupervised Personality Recognition for Social Network Sites (Celli, 2012)	linguistic differences on personality (Big 5), USML: score -based, 748 Italian FriendFeed users (1065 posts)	good reliability of the personality model, the most frequent personality type is: extravert, insecure, agreeable, organised and with low openness	personality is treated as binary variable

Table 2.4. Comparison of different methods of determining personality profile based on various data - part 2

No	Paper name	Techniques	Advantages	Limitations
6	Predicting Personality Using Novel Mobile Phone-Based Metrics (Montjoye et al., 2013)	Big 5 derived from Standard Mobile Phone Logs (CDR). Published Correlation co-efficient.	SVM method for class prediction, average accuracy 42%, banchmarked to random; the best predicted traits were Extraversion and Neuroticism	CDR is data that is particularly protected by law and is not available without special consent from users; very small sample N=69, no evidence about validation
7	Personality, gender, and age in the language of social media: The open- vocabulary approach (Schwartz et al., 2013)	Big 5 derived from Facebook posts (77,000) using new open-vocabulary technique and LIWC as baseline, personality was discrete binary outcome	pioneer research using the open vocabulary approach, accuracy: gender-0.919, age-0.84, E-0.38, A-0.31, C-0.35, N-0.31, O-0.42	personality as a binary class
8	Multimodal Classification of Personality States In The Wild (Kalimeri et al., 2013)	data from sociometric badges (accelerometer, microphone, Bluetooth, infrared sensors)+e-mail data. SVM ML algorithm + feature selection (sequential floating search))	average accuracy ratio from 0.45 - 0.57) the highest for Emotional Stability trait and Openness	6 weeks period of observation using specific not commonly used (yet) equipment; very small sample N=54, no evidence about validation
9	Personality Traits Recognition on Social Network- FB (Alam et al., 2013)	Big 5 derived from text (FB) SML, N=250	baseline study for personality recognition with macro-averaged precision=0.58, recall=0.59, F1=0.59, accuracy=0.62	personality was treated as binary variable
10	A Comparative Study of Different Classifiers for Myers-Brig Personality Prediction Model (Chaudhary et al., 2013)	MBTI model derived from online texting, SML techniques	the ML algorithm improve the accuracy by tuning its parameters from 0.6659 to 0.6775	MBTI is a 4 binary features typology

Table 2.5. Comparison of different methods of determining personality profile based on various data - part 3

No	Paper name	Techniques	Advantages	Limitations
11	Automatic Personality Assessment through Social Media Language. Journal of Personality and Social Psychology (Park et al., 2014)	MyPersonality users (71,556), Big 5 predicted from tweets, regression predictive model, LDA, RPCA	complex but well documented research, multi-staged analysis with final validation from re-test, correlation between prediction and test result was the highest for O-0.43,E-0.42, C-0.37, A-0.35, N-0.35	the analytic sample was limited to users who wrote at least 1,000 words (status messages), shared their gender and age (less than 65)
12	Predicting personality on social media with semi-supervised learning (Nie et al., 2014)	Big 5 on Facebook MyPersonality App data and publicly available profile data	it works with group of texts, rather than a single text, and does not relay on users profiles and achieved accuracy is 0.83 for some traits	automatic analysis is difficult due to colloquial language; no evidence about validation
13	The Online Social Self: An Open Vocabulary Approach to Personality (Kern, Eichstaedt, Schwartz, Dziurzynski, et al., 2014)	Data from 69792 FB users, via MyPersonality app, correlation between personality traits (Big 5) and online word usage	using LIWC to investigate correlation between personality and the language used by people; the method of correlations of words categories is more efficient than individual words and phrases	no predictive model was built based on that research. The results are limited to FB users
14	What Your Face Vlogs About.. (Teijeiro-Mosquera et al., 2015)	Expression Recognition Toolbox (CERT) System to characterise users of conversation vlogs	the common sense knowledge with sentiment information and affective labels increased the accuracy of the existing frameworks	using CERT dictionary alone to analyse the data is not sufficient; it also result in a very small number of feature extractions.

(Source: own work)

Table 2.6. Comparison of different methods of determining personality profile based on various data - part 4

No	Paper name	Techniques	Advantages	Limitations
15	Personality classification based on Twitter text using NB, KNN, SVM (Pratama and Sarno, 2015)	Big Five based on text from Twitter, ML, 10.000 status updates from 250 users (My personality data set), modified into one long string (250 documents)	the results are predictions for each Big Five traits, primary personality profile and secondary personality characteristics which were obtained from the combination between two traits, achieved accuracy 0.6	multi-label method used in this experiment is a binary relevance that is transforming each label in binary with independent assumption
16	Personality Traits on Twitter-or-How to Get 1,500 Personality Tests in a Week (Plank and Hovy, 2015)	text (Twitter), corpus of 1.2M English tweets annotated with MBTI personality, ML	analyze which features are predictive of which personality traits type and gender, achieved the best accuracy is for Stability and Extraversion	results shows the important bias of gender in model accuracy
17	Age Gender and Personality recognition using tweets in multilingual settings (Arroju et al., 2015)	ML,multi-linguistic model for Big 5 personality, gender and age using LIWC and ERCC, c.a 150 users of Twitter	integration of many tools for automatic analytics, achieved accuracy=0.68	small sample and unique method of performance evaluation do not allow to make any evaluation
18	Identify Website Personality, by using Unsupervised Learning Based on Quantitative Website Elements(Chishti et al., 2015)	Big 5 prediction based on the quantifiable website content, USML	model can be used for automatic personalising of a website, Accuracy score from K-mean is K=10	the system highly depends on the user's activity on the website and can be enhanced by integration with data from other websites
19	TWISTY: a multilingual twitter stylometry corpus for gender and personality profiling (Verhoeven et al., 2016)	personality (MBTI) derived from multi-lingual text (6 languages), SML	the method outperformed for gender, Extraversion and Stability	Personality as a set of 4 binary features

(Source: own work)

Table 2.7. Comparison of different methods of determining personality profile based on various data - part 5

No	Paper name	Techniques	Advantages	Limitations
20	A comparative Study of different Classifiers for automatic personality prediction (Ngatirin et al., 2016)	students Extraversion predicted on the basis of text, WEKA platform is used, LIWC and SML techniques	the research purpose was to make a comparison of different ML algorithms, based on F1 score, the highest was OneR 0.84	only one trait was studied in the experiment and the cases were very homogeneous concerning age, a small sample N=100
21	Social Media User personality classification using computational linguistic (Lukito et al., 2016)	MBTI based on Twitter texts, WEKA software, S-SML	not only the linguistic parameters were analysed, but also the features connected with behaviour, re-activeness, elements used in tweets, good performance only for Extraversion	only account with more than 500 Tweets were analysed
22	Analysing Personality through Social Media Profile Picture Choice (Liu et al., 2016)	Big Five based on text and correlated with profile picture features, twitter users, r-Pearson correlation and regression analysis	large scale research (66,000) using personality derived from data and investigating the correlation with a wide range of interpretable aesthetic and facial features	the article referred to the method presented in (Schwartz et al., 2013) and (Park et al., 2014) and presents only the correlation matrix between traits and groups of features
23	Personality prediction based on Twitter information in Bahasa Indonesia (Ong, Rahmanto, Suhartono, et al., 2017)	Big 5 derived from Twitter text in Indonesian, 359 of users, SML	analytic was performed on new collected data with using SVM and XGBoost	personality was evaluated by 3 experts who labeled tweets with binary labels (high or low) for each of the Big 5 traits, none psychometric tool was used and the outperformed accuracy level (0.97 for XGBoost) are probably the result of evaluating mistake (CV procedure treated as a evaluation method)

Table 2.8. Comparison of different methods of determining personality profile based on various data - part 6

No	Paper name	Techniques	Advantages	Limitations
24	Towards User Personality Profiling from Multiple Social Network (Buraya et al., 2017)	MBTI model based on multi-source text data set (NUS-MSS), SML	multi-modal data fro, m three social networks (Twitter, Foursquare, and Instagram), the early fusion of multi-source data helps to improve classification performance by more than 17% in some cases	the article does not contain relevant data allowing to assess the correctness of the study
25	25 Tweets to know you: A new Model to Predict Personality from Social Media (Arnoux et al., 2017)	Big 5 based on text, N=1323, Word Embedding with Gaussian Processes, USML	this method outperforms the other methods for 37% for 200 tweets, for 25 tweets it is performing on par with the state-of-the-art, personality results were normalised to 0-1	the method achieved an average correlation of 0.33, with the best performance for Neuroticism at 0.42
26	Persona Traits Identification based on MBTI - A Text Classification Approach (Bharadwaj et al., 2018)	personality based on SM online texting, EmoSenticNet, LIWC	a complex approach of investigating personality together with emotions, the results are displayed on Web Application	a huge data set was used, an average number of tweets per user: 49, average number of words per user: 1310, it's not clear how the probability of a given trait is calculated based on a binary class
27	Reddit: A Gold Mine for Personality Prediction (Gjurković and Šnajder, 2018)	MBTI model was used on reddit posts (open data MBTI9K dataset), LIWC, SML	MLP has the best performance	the model predicts the type of personality, each trait is predicted separately as binary class

(Source: own work)

Table 2.9. Comparison of different methods of determining personality profile based on various data - part 7

No	Paper name	Techniques	Advantages	Limitations
28	Facial-Based Personality Prediction Models For Estimating Individuals Private Traits (Bin Tareaf et al., 2019)	Social Media Photo was analysed by Face ++ API and linked with personality prediction made by Facebook My Personality App. Pearson Correlations and ML techniques (SVR, RF, BA, AB)	50 facial features are enough for predicting Big 5 personality profile with MSE ranged 0.05 for Extraversion to 0.09 for Openness	the use of social media data limits the possibility of using this method to personalise other services; profile photos do not have to be photos of the account owner
29	Detecting Personality Using Eye-Tracking Data(Berkovsky et al., 2019)	using eye-tracking data and features for predicting Hexaco personality model; ML technique (AB, DT, LR, NB, RF, SVM, kNN) N=18	high average accuracy 0.85 probably resulting from the homogeneity of the data sample N = 18	method to be used as a laboratory test in a psychological study or for VR games personalisation
30	Personality Prediction from Social Media Images: A Content Driven Approach (Sahu et al., 2019)	CNN Deep Neural Network is used to classify the objects published in SM photos, and based on that the interest of the person is deducted	this technique can be used for personality prediction but needs to be analysed with the personality classification	the technique is hermetic, probably limited to the social media data owner (authors analyse their own photos only)
31	Predicting personality from patters of behaviour collected with smartphones (Stachl et al., 2020)	ML (not specified with cross validation), based on 30 days data from smartphone usage and Big 5 60 item tool, N=624 volunteers from university groups, r-Pearson correlation reported only	the links between smartphone behaviours and 30 facets of Big 5 dimensions gives strong evidence for creating a model without data linked with the Big 5 results on the smartphone data	30 days of observation are hard to achieve, privacy issues, group of volunteers highly educated mostly

(Source: own work)

Table 2.10. Comparison of different methods of determining personality profile based on various data - part 8

No	Paper name	Techniques	Advantages	Limitations
32	Personality Classification from Online Text using Machine Learning Approach (Khan et al., 2020)	Emolex, LIWC and TF/IDF for vectorisation of English language text, ML Techniques SVM and XGBoost, and Myer-Briggs Type Indicator (MBTI) model; data set is taken from Reddit/Kaggle; personality derived from My personality App	application of re-sampling technique for class imbalance problem very high parameters of the models were achieved (Accuracy in range 60-99.92 (XGBoost)	Personality is a binary classification for 4 MBIT traits; recruitment from Twitter forum there is no information what amount of text per person is required to achieve good prediction
33	Classification of Proactive Personality: Text Mining Based on Weibo text and Short-Answer Question Text (P. Wang et al., 2020)	ML methods (SVM, XGBoost, KNN,NB, LR) with cross validation, the best results achieved with SVM and LR, N=901 and analysis was taken on 4955 posts in total, personality was evaluated based on 4 open-ended questions giving binary results of pro-active personality (50% of participants)	feature extraction can improve the accuracy of prediction	material from social media and artificial definition of personality (one dimension of pro-activeness)
34	PANDORA talks: Personality and demographics on Reddit (Gjurković et al., 2020)	Big 5 using MBTI binary class as independent variable, Reddit posts, SML	an unique study combining the MBTI typology, the Big 5 and another Enneagram model in one experiment; MBTI and Enneagrams labels were used for predicting Big 5 trait	poor performance of deep learning baseline model, further efforts should be focused on improving the sample representativeness, larger samples of materials from users and deep ML procedures and architecture

(Source: own work)

Table 2.11. Comparison of different methods of determining personality profile based on various data - part 9

No	Paper name	Techniques	Advantages	Limitations
35	Multitask learning for Emotion and Personality Detection (Y. Li et al., 2021)	training model is built for emotion and personality together; text methods; a dataset corpus for personality contains 9917 multi-labeled (Big 5 dimensions) sentences ; ML models are SiGMTL, CAGMTL, SoGMTL, and new-created MAML method; personality dimensions are treated as binary variables,	authors proposed some methods of increasing the prediction accuracy and recall	personality is treated as a typology of five binary dimensions, there is no metrics for separate traits

(Source: own work)

The 35 studies presented in the tables (Tab. 2.3, 2.4, 2.5, 2.6 2.7, 2.8, 2.9, 2.10, 2.11) are not the only ones published. The research was selected according to the popularity of citations, with particular emphasis on methods based on machine learning techniques. Based on just this sample of 35 studies, it can be concluded that:

- The basic data on which personality is determined is text data (27 out of 35 studies) from posts and tweets on social media;
- Social media is also the most common source of data for determining the user's personality (25 out of 35 studies). This is because of the availability of open data collections from the MyPersonality App, which many data scientists use. They are also the basis for research work, and it improves the methods of machine learning concerning classification problems of imbalanced samples. However, while this data has some value from the point of view of improving the methods themselves, using them has little value in personalising services by third parties other than social media. This data is utilised mainly for business purposes by providers of social-media portals (e.g., Facebook, Twitter, Weibo). The popularity of analysis is related to the availability of ready-made sets marked with personalities or emotions on open-source resources from dedicated websites (e.g., Reddit, Kaggle). Most publications written based on analysis of open-source data are to search for

the best analytical method. No cases were found that describes the application of such a model for business purposes other than adding personalisation on Twitter and Facebook;

- In addition to the text, additional type of data taken from social media are: profile photos or photos published in social media (3 of the 35 studies presented);
- There are single attempts to build models predicting user personality on other data such as CDR from telco-operator, socio-metric badges, eye-tracking equipment, data collected from 30 days of smartphone usage;
- For two years machine learning, has been the dominant method of building models predicting personality. (9 of the 17, and 100% if consider those after the year 2019). Based on the ten-year history of research on personality determination from data, machine learning is the dominant method (26 of the 35 studies).
- There are two dominant personality models used in the researches: Big 5 (Costa and McCrae, 1992) described in Section 2.2.2 of this dissertation and Myers-Briggs model (MBTI) (Isabel Myers, 2010) briefly presented in Section 2.2 of this dissertation. The Big 5 is often used (21 of the 35), while the MBTI is used at least twice less frequently (9 of the 35). The other 5 are single uses of other models. Regardless of the model used (Big 5 or MBTI), in most cases, it can be stated that the personality model is treated as a set of discrete binary variables. This approach is coherent with the Myer-Brigs Theory, which in fact is a typology composed of a combination of 2-pole classes, which gives 16 personality types. However, the use of Big 5 dimensions in predictive models as a binary classifier is not consistent with interpreting the score in psychometric Big 5 inventories and the Big 5 Theory of traits.
- Moreover, although it is convenient from the point of view of building models (it avoids imbalanced samples), it does not seem justified from the point of view of personalising products, where the idea of personalisation is to distinguish people with atypical needs. For personalisation, personality traits should be used to identify those who differ significantly from the typical, average level of the trait. The Big 5 model makes it possible because traits are dimensions, and it is possible to freely define the center of the population (typical users) and the cut-off points of extreme groups (groups with special needs). However, this fundamental

difference affects both the interpretation and possible use of the result itself, and causes problems with prediction (Stajner and Yenikent, 2021) - binary typology is less likely to differentiate behaviour.

- What is important is that regardless of the model used (Big 5 (21) or MBTI (9), in most cases (in the case of Big 5, 17 out of 21 cases), it can be stated that it is a discrete binary variable. Therefore, it is used in predictive models as a binary classifier. What, particularly in the case of the Big 5, is not consistent with the interpretation of psychometric tools for measuring personality, nor does it seem justified from the point of view of personalising products. Details on a different approach to personality as a classifier and the reason are described in Section 2.4.1.

The increase in the importance of machine learning methods shown in the review is a direct consequence of the rapid development and popularisation of artificial intelligence construction techniques. As a result, every year, the availability of more and more ready-made and easy-to-use tools enabling various machine learning analysis is increasing. As these methods are used in this work and they are discussed separately in more detail in the next chapters.

2.3.2 Quality Criteria Concerning Building Personality Models Based on Data

The systematic literature review resulted in meta-analysis made by (Azucar et al., 2018). In this extensive review and discussion about documented and proven examples of determining the user's personality based on his/her digital footprint, the researchers compared 26 studies, 16 of which were considered by authors as ethical, complete and reliable. The comparative analysis has driven to defining the universal and essential criteria allowing for evaluating the research of personality determination based on data. Authors conclude that in the early 2010s, most of the research focused on social media data, and most studies used traditional statistical methods such as Pearson's r-correlations, statistical predictive models or machine learning-based predictive models. Although in recent years, in line with the predictions of the authors (Azucar et al., 2018), there has been a constant transformation from traditional analytic to modern analytical methods: data mining and machine learning algorithms. There are a few examples of improving the quality of these predictions, but it must be taken into account that the authors limited themselves to research using SM data (FB, Twitter) and expecting the analysed data categories list expansion.

In the mentioned review of research (Azucar et al., 2018) on personality prediction from digital traces, the following classification criteria concerning building personality models based on the adopted data.

1. In the study, digital data had to be linked to personality scores at the individual (personal) level.
2. Personality should be measured based on a standardised and validated personality measurement tool. However, it did not necessarily have to be a diagnostic tool (the list of used tools also includes 10-items tools - which means that each personality dimension was evaluated based on two statements).
3. Digital data had to be collected automatically.

The fulfilment of these three criteria was considered necessary for the research to be considered valuable. Therefore, the method planned for this dissertation that determines the user's personality based on digital footprint should fulfil all three criteria. The only difference in this study is the data type. Instead of various data recorded from social media (mainly texts), in research for dissertation, various smartphone data was used (data characteristic collected and finally used for this dissertation is described in Section 3.5).

2.3.3 General Information about Machine Learning

Currently, ML methods are commonly used to solve problems, e.g. classification, regardless of the field of analysis (Raschka, 2018). In the Figure 2.1 there is presented general scheme of individual steps inside the data analytic pipeline when executing its fit method on the training data and the predict method on the test data. Raschka (Fig. 2.1) compares creating a class prediction model to an analytical pipe where raw data is set as input and a model predicting the class label is built through a sequence of analytical transformations. These transformations rely on repeatedly transforming and fitting the output variables through scaling, dimension reduction, training the class prediction

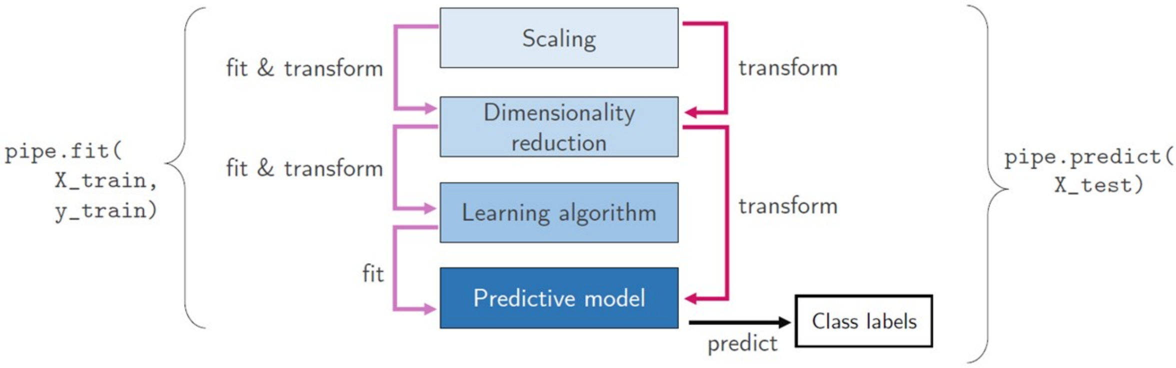


Figure 2.1. Scheme of individual steps inside the pipeline: from Raschka et al. (2020, p.7)

algorithm, and finally fitting all elements in the predictive model. As a result, a model is created that predicts the desired class with some accuracy based on other, not directly related, data.

2.3.4 Performance Evaluation Metrics for Classification Problems

In this section, performance evaluation metrics of predictive models from published studies is presented. Unfortunately, not every published study provides complete information on the quality and parameters of predictive models. First of all, general measures for predictive models built with the machine learning method are discussed. The second section presents the comparison of metrics from existing researches related to personality prediction.

2.3.4.1 Performance Evaluation Metrics related to Machine Learning Methods

One of the most popular and widely used metrics available in Machine Learning software is Confusion Matrix. Confusion Matrix is used to evaluate the validity and accuracy of the model. It is most commonly used when there is a classification problem where the output may contain two or more class types (Sokolova and Lapalme, 2009). The Confusion Matrix itself is not a measure of performance, but it does contain the data on which nearly all performance metrics are based. The four basic measures are: true positive (TP), true negative (TN), false positive (FP), and false-negative (FN). TP is a measure of correctly identified (predicted) classes. Similarly, TN is a measure of correctly identified cases of non-belonging to a class. Therefore, FP is a measure of incorrectly assigned classes for cases that did not belong to a class, and FN refers to assigning a non-belonging to a class in a situation wherein fact, the case belongs to that class. The Table 2.12 shows the interpretation and dependence of these measures for the binary classification: With these definitions in mind, the main goal is to maximise

Table 2.12. The Confusion Matrix for Binary Classification Prediction

	real positive class (1)	real negative class (0)
predicted positive class (1)	TP	FP
predictive negative class (0)	FN	TN

the diagonal measures from the upper left corner of the matrix to the lower right corner of the matrix to search for the best predictive model. A model that ideally predicts cases should give 0 false positives and 0 false negatives. In reality, however, no model will be 100% accurate in most cases. Based on the measures from the Confusion Matrix, other performance measures can be constructed. Those that are common in practice covers (Sokolova and Lapalme, 2009; Visa et al., 2011): accuracy, precision, recall (sensitivity), specificity, and F1 score.

$$Precision = \frac{TP}{TP + FP}$$

The *precision (or positive predictive value)* is the proportion of correctly predicted positive]objects to the number of all objects predicted correctly.

$$Recall = \frac{TP}{TP + FN}$$

The *recall* (or sensitivity or true positive rate) is the proportion of correctly predicted positive objects to the number of all positive objects (TP+FN).

$$Accuracy = \frac{TP + TN}{n}$$

The *accuracy* is the proportion of correctly classified in prediction (TP+TN) to the number of all objects from data set (n).

$$F1score = 2 * \frac{precision * recall}{precision + recall}$$

The *F1 score* (or F-measure) is a harmonic mean of recall and precision but scaled to 1 (in extreme situation recall and precision can be equal to 1).

$$ErrorRate = \frac{FP + FN}{n}$$

The *error rate* is the proportion of all incorrectly assigned objects to the total number of all objects (n). Can be treated as a average classification error.

$$Specificity = \frac{TN}{FP + TN}$$

The *specificity* (or true negative rate) is the proportion of correctly predicted negative objects to the number of all negative objects (FP+TN). These basic measures are used to evaluate the model from different perspectives. Therefore, not all of them have to be taken into account in the evaluation of predictions. In the machine learning process for multi-class prediction and choosing the best method, it is impossible to maximise all measures simultaneously. However, when assessing a model itself and formulating qualitative criteria for its assessment, one should always consider its subsequent model and the assumptions made. The model's business goals will not always be the same as the overall goals for machine learning where *Accuracy* is maximised. Such a problem exists in the case of strongly imbalanced class predictions where it is easy to obtain a high average *accuracy* for the entire model by making predictions for the most numerous class. The possibility of resolving the imbalanced class problem will be discussed in more detail on the example of the finished model.

2.3.4.2 Performance Evaluation Metrics related to Personality Prediction Machine Learning Models

In this Section, examples of personality prediction using the methods of machine learning (Supervised Machine Learning (SML), semi-supervising (S-SML), and unsupervised

(USML) can be found. In addition, the Tables 2.13, 2.14, 2.15 contain information about the analysed data, methods, and performance measures if presented in the publication.

Table 2.13. Comparison of different performance measures related to personality predictive models

No	Paper	Data Category	ML Method	Performance
1	Celli (2011)	Big Five from Tweets individually written	SML	mean accuracy=0.665
2	Golbeck et al. (2011)	Big Five from Tweets	SML: regression models: ZeroR and GP	Accuracy range: 0.42 (N) to 0.75 (O)
3	Quercia et al. (2011)	Big Five from Tweets	SML: M5 rules with 10-fold CV	RMSE: 0.69 (O),0.76 (C), 0.79 (A), 0.85 (N), 0.88 (E)
4	Celli and Rossi (2012)	linguistic differences on personality	USML model valuated by experts	coherence between model assessment and expert assessment 0.62
5	Celli (2012)	linguistic differences on personality (Big 5)	SML: score -based	Accuracy of model 0.63
6	Alam et al. (2013)	Big Five from FB text	SML: MNB performs better than SMO and BLR[2]	MNB=0.62, BLR=0.57, SMO=0.58
7	Chaudhary et al. (2013)	MBTI model from online texting	SML: NB, SVM, LR, Random Forest	Accuracy:LR=0.66, SVM=0.65, NB=0.56
8	Kalimeri et al. (2013)	data from sociometric badges (accelerometer, microphone, bluetooth, infrared sensors)+e-mail data	SML: SVM + feature sequential floating search	average accuracy ratio from 0.45 - 0.57, the highest for Emotional Stability trait and Openness
9	Pratama and Sarno (2015)	Big Five from Tweets	SML: KNN, NB, SVM	Accuracy: KNN(0.58) NB(0.6) SVM(0.59)

(Source: own work)

[2]SMO - Sequential Minimal optimisation for Support Vector Machine, BLR - Bayesian Logistic Regression, MNB - Multinomial Naïve Bayes

Table 2.14. Comparison of different performance measures related to personality predictive models

No	Paper	Data Category	ML Method	Performance
10	Plank and Hovy (2015)	MBTI model, text (Twitter)	SML: Logistic Regression	Accuracy for traits: E=0.72, S=0.77, T=0.61, J=0.55
11	Arroju et al. (2015)	linguistic model for personality, age and gender based on tweets	S-SML: SGD classifier, LIWC with ERCC (regressor model)	Accuracy=0.68
12	(Chishti et al., 2015)	Big 5 model based on elements of website content	USML: K-mean	Accurate score K=10
13	Verhoeven et al. (2016)	gender and personality (MBTI) from tweets corpus (Twisty)	SML: SVM, LR	F1 scores: E=0.77, S=0.79, T=0.52, J=0.47
14	Ngatirin et al. (2016)	students personality based on twitter	SML: NB, SMO, RF, AdaBoostM1, OneR	F1 score (OneR) 0.83
15	Lukito et al. (2016)	MBTI based on Twitter	S-SML:lexicon and linguistic rules based	Accuracy 0.8 for E, other dimensions 0.6
16	Ong, Rahmanto, Williem, et al. (2017)	Big 5 for tweets	SML: XGB, SVM	Accuracy XGB=0.97, SVM=0.76
17	Buraya et al. (2017)	multi-source data set (NUS-MSS)	SML: not specified	internal baseline, after vectorisation of all data sources to one vector 17% improvement was achieved
18	Arnoux et al. (2017)	big 5 on Tweeter usage	USML: word-embeding	Accuracy 0.68
19	Bharadwaj et al. (2018)	MBTI personality on online text	TF-IDF, LIWC, SML: SVM, NN, NB	the best Accuracy for SVM: E/I(0.85) S/N(0.88) T/F(0.87) J/P(0.79)

(Source: own work)

Table 2.15. Comparison of different performance measures related to personality predictive models

No	Paper	Data Category	ML Method	Performance
20	Gjurković and Šnajder (2018)	MBTI, reddit posts	SML: SVN, LR, MLP	MLP has the best performance for (accuracy): E/I=0.83, S/N=0.79 and LR for: T/F=0.67, J/P=0.75m
21	Bin Tareaf et al. (2019)	Social Media Photo with FB My Personality App.	SML techniques (SVR, RF, BA, AB)	MSE ranged from 0.05 for Extraversion to 0.09 for Openness
22	Berkovsky et al. (2019)	eye-tracking data and features for predicting Hexaco personality model	SML technique (AB, DT, LR, NB, RF, SVM, kNN)	high average Accuracy 0.85
23	Sahu et al. (2019)	SM photos	CNN Deep Neural	not specified
24	Khan et al. (2020)	MBTI on dataset from Kaggle resources from social network	SML XGBoost	Accuracy 0.95-0.99
25	Stachl et al. (2020)	30 days data from smartphone usage and Big 5 60 item	ML (not specified with cross validation)	r-Pearson correlation for features reported only
26	Khan et al. (2020)	texiting from social media (open data Reddit|Kaggle), MBTI model (binary)	Emolex, LIWC and TF|IDF for vectorisation, SML Techniques SVM and XGBoost	XGBoost Accuracy 0.60-0.99
27	P. Wang et al. (2020)	Social media post, personality assessment based on 4 open questions	SML: SVM, XG-Boost, KNN,NB, LR with cross validation	, the best results achieved with SVM and LR
28	Gjurković et al. (2020)	Big 5 prediction, using gender class and MBTI binary class as independent variable, based on reddit posts	SML: LR	Regression (Pearson correlation coefficient): E(0.39), A(0.28), C(0.27), N(0.28), O(.26)
29	Y. Li et al. (2021)	texts based and Big 5 (binary)	a set of novel multitask ML based on CNN	highest accuracy achieved among experiments was 0.63, range:0.60-0.63

Unfortunately, not all published studies contain metrics and details allowing for comparison of the predictions efficacy. The above list was intended to present the metrics (Accuracy, F1 score) that are available and can be used as a state-of-the-art benchmark. The set of metrics for this benchmark is presented in the Figure 2.2. The range of the prediction accuracy of models has wide margin and vary from 0.45 to 0.85 (2.13, 2.14, 2.15. The highest score is for a single Big 5 computation study from the eye-tracking register. It can be concluded that this is entirely consistent with the biological (temperamental) conditions of the personality in the Big5 theory. The eyes are an inherent part of the nervous system, hence the high accuracy of prediction. It can even be hypothesised that if it were not for the measurement error associated with the Big 5 inventories themselves (declarative behaviour assessments), this result would be even higher.

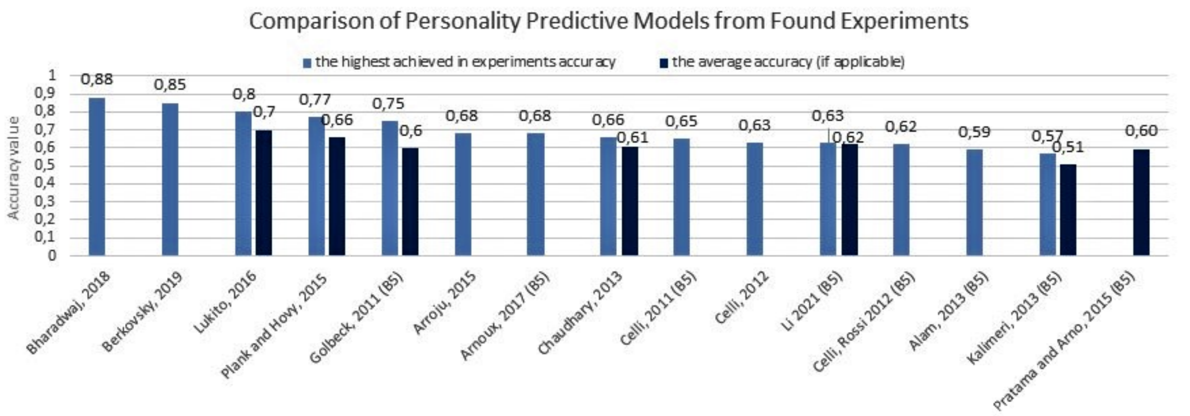


Figure 2.2. Comparison of Accuracy values from found publications. Source of the numbers is given as citation on X axis. Additionally the models predicted Big 5 personality are marked by B5)

(Source: own work)

Out of 29 reported publications, 3 were not considered as a valuable benchmark (In 2.14 number 16, and in 2.15 numbers 23 and 25 (both of the same researcher). The validation process is vaguely described and the reported results indicate model fit accuracy rates on sets intended for training, rather than real evaluation on the separate sets not used for training. The highest accuracy among the other Big Five prediction models was reported by Golbeck et al. (2011)- 0.75, and this is the result achieved using prediction from regression models by comparing to real value on extremely small sample of 50 Twitter users, and on the base of 2000 tweets per user. The remaining studies based on SM texts and the Big5 model reported the highest achieved accuracy of the models in the range of 0.51-0.68. Finally, it can be concluded that models based on text data achieved higher accuracy than the models built based on data collected from socio-metric badges (the mean highest accuracy for text data is 0.68 and socio-metric badges 0.57). Considering the original assumption of Big 5 theory described in Section 2.2.1 it can be concluded that lexical differences should naturally be a better source for personality prediction than other indirect data like the way of using a smartphone. But unfortunately this kind of data are not available during service installation.

Even though, as a result of the literature review, it was impossible to find fully comparable studies in terms of the set of data used for prediction and three-class personality profiles. Some studies will serve as a benchmark in evaluating the created models. First of all, the purpose of these predictions is the same, which allows for an approximate assessment of the general reference level for predicting similar aspects in psychology and social research.

2.4 The Value of Knowledge About Personality in Contemporary Enterprise and Service Creation

In this chapter, various aspects of the value of knowledge about personality of electronic services users in a modern enterprise will be discussed. It seems that the popularity of research on the automatic determination of personality from data (discussed in Section 2.3.1) and numerous attempts to build communication with the client (advertising) on this basis prove the attractiveness of this solution compared to traditional methods of building knowledge about the client.

2.4.1 Personality as a Classifier of Service User Needs

Matz et al. (2017) confirmed that the use of psychological targeting allows influencing the behaviour of social media users by adapting the message to the psychological needs of recipients defined by personality (Big 5). However, at the beginning of using a new type of AI service, such as a virtual/voice assistant, the data from the history of usage will not be available. In the context of even better regulations on protecting users' privacy, it seems inevitable to achieve the point in which services will be limited to the use of their own data only, and data information systems will become more and more hermetic. Therefore, there is a need for creating methods that will allow services to detect user's personality based on even small amounts of data available at the time of service installation, e.g., from a mobile phone (apps used, history of usage, notes, calendar, photos, way of data storage). In this paradigm, the assumption refers to determining the personality of each user of a smart device (personal or home), so the model should be based on the data available on different smart devices. Therefore, such data types that do not apply to everyone (e.g., social media or other specific services or applications) were excluded from investigations.

Until now, the user's personality has been measured for psychological diagnosis by standardised and normalised psychometric tools (Magnusson, 1967). These tools are standardised for a specific language version and a specific population and population subgroups, usually described by gender and age. Therefore, the raw score is not interpretable and requires reference to the population. Additionally, according to the Big 5 model, these features are never binary in personality dimensions. The differentiation between people must be statistically significant. According to the statistical properties of the normal distribution, the closer to the centre of the normal distribution, the greater the difference between the unit results must exist to be considered statistically significantly different.

The literature review presented shows that in most research on the automatic determination of personality, dimensions are incorrectly treated as binary variables. By assumption of psychological theory, there is an observable behavioural continuum and, by definition, dimensional range cannot be considered as binary variable. Adopting such an assumption in most of the existing studies is convenient from the analytical point of view because it eliminates class imbalances in the analyses. However, this approach cannot be applied when personality is used to predict user needs for service personalisation.

In this application, the challenge is to identify people with special needs. Personalisation is used to meet the needs of atypical people; typical people are usually the recipient of a specific product while defining which one is looking for the most common needs. Describing it in the statistical language, this typical recipient is about 68% of the population, falling in the range from 1SD to + 1SD from the average.

In this respect, the individuals do not differ significantly. However, as a consequence of the binary subsection into two classes according to the median, it classifies typical people (1SD from the mean) and atypical people, i.e., 1SD and 2SD average, into one class. If the characterisation of personality from digital data is to replace traditional questionnaire methods, the nature of the personality dimension cannot be changed and replaced with a binary nominal variable. To sum up, both, from the point of view of the theory of personality traits and from the point of view of using personality dimensions as a classifier of needs and the basis for personalisation, the method must enable the identification of people with atypical needs (taking into account the significance of this atypicality at the individual level).

2.4.2 Role of Personality in Human - Service Interaction

Each subsequent technology and subsequent intelligent services will be increasingly complex and challenging to learn. It seems that the adaptation process will be more and more difficult. Understanding the user's personality at the beginning of the adaptation process seems crucial for applying an appropriate activity to eliminate cognitive barriers. The adaptation process is easier for high Extraversion, high Conscientiousness, and low Neuroticism (Matthews, 2008). Highly neurotic people are generally not resistant to stress, accompanied by a higher level of anxiety and a lower ability to adapt, which stands in contrast to high Openness, as it is a beneficial feature when learning about and discovering new things. An additional incentive for people with high Openness to using such technologies is their intellectual involvement.

According to Matthews (2020), robots and virtual agents simulate human behaviour better and better to facilitate contact and cooperation (rich communication, support, kindness, warm and deep contact). Even simulating human reactiveness promotes the anthropomorphisation of virtualised entities. Interestingly, people with high Extraversion and Emotional Stability (low Neuroticism) anthropomorphise robots and virtual agents more often. They create emotional closeness, which facilitates contact and cooperation (Salem et al., 2015). This dependency also works the other way round; robots with highly extroverted behaviours give a more favourable impression than robots with introverted features (Robert, 2018), which can be a good solution in the absence of information about the user's preference. In turn, research by Q. Zhang et al. (2019) shows that mapping a user's personality can facilitate contact with robots or other intelligence simulations. Extroverts prefer robots with extroverted behaviours and introverts with introverted ones.

In addition, more and more such research is carried out in the context of human-AI interactions (no matter whether in a virtual form or in the form of a physical robot). These studies often focus on social media use and problematic or dysfunctional internet use (e.g., N.S. et al. (2019) and Stead and Bibby (2017)). These studies go far beyond the scope of this study and do not directly concern the automation of adjusting the operation of an intelligent service to the user's needs. For example, research on Facebook concerns the personalisation of advertising content. Due to the purpose of this dissertation, the review will be limited to interactions with advanced robots and virtual agents. According to Robert (2018), three areas of research into the personality of advanced robots can be distinguished:

- Influence of human personality on the manner and quality of human-robot interaction (HRI).
- Influence of the simulated robot's personality on the way and quality of HRI.
- Influence of the match between human personality and robot personality on HRI. Here, the main model studied is the similarity of human and robot personality.

The conclusions of Robert's research can be extrapolated and generalised to human interactions - advanced AI-based services - even if these are purely virtual services. The participant's comfort rating plays a crucial role in assessing the quality of these interactions. Thus, it was discovered that people with high extraversion and emotional stability anthropomorphise robots and create psychological relationships more often than others and with greater naturalness (Salem et al., 2015). Additionally, they feel much more comfortable than the rest in such relationships (Syrdal et al., 2006). However, these close feelings do not necessarily imply a higher level of trust. Extroverts tend to prefer extroverted robots, as opposed to introverts (Robert, 2018). However, people prefer to have opposite personalities in some contexts, e.g. automatic cars (Q. Zhang et al., 2019).

2.5 Personalising Interactive Services

The purpose of this dissertation is to use the user's personality profile to personalise any e-service or mobile service. This and the following chapters are devoted to personalisation methods, connecting personalisation with new technologies and their importance in modern business. personalisation can be defined broadly and in many different ways. Here are two examples of definitions:

"personalisation is about building customer loyalty by building a meaningful one-to-one relationship; by understanding the needs of each person and helping to achieve a goal that effectively and competently addresses the needs of each person in a given context" (Riecken, 2000).

'personalisation is the ability to deliver content and services tailored to people based on knowledge of their preferences and behaviour' (Hagen et al., 1999).

The first definition is related to creating and maintaining relationships with customers. The second relates to the specific activity of delivering matched content. According to Fan and Poole (2006), both research and business practice focus on how to personalise, and less attention is paid to quality assessment and effectiveness of personalisation itself. Studies comparing the efficacy of personalisation are rare. The lack of consensus as to the definition of personalisation is also evidenced by the fact that the expression is used interchangeably with other terms: customisation, adaptation, individualisation, one-to-one relationship and finally customer-centric approach. Ball and other researchers link personalisation with customer satisfaction and loyalty improvement and churn (Ball et al., 2006).

Max Weber, half a century ago described the perspective of personalisation in the ideal type theory(Wagner and Rogers, 1969). Weber argued in his works that social and economic research is conceptual in nature and will never be fully inductive or descriptive. Webers' ideal type theory, which is an abstraction of the fundamental characteristics of a specific community or economy, is helpful in the study of personalisation (Fan and Poole, 2006). Researchers identified four ideal types of personalisation: architectural, relational, instrumental and commercial perspectives (Table 2.16). Each perspective represents a different motivation behind personalisation, defined by the goal, strategy and techniques of modelling the user of the service. The characteristics of each of the ideal types also result in different criteria for assessing personalisation.

Table 2.16. Personalisation Ideal Types from (Fan and Poole, 2006), page

Architectural	**Instrumental**
Motive: To fulfil a human being's needs for expressing himself/herself through the design of the built environment	**Motive**: To fulfil a human being's needs for efficiency and productivity
Goals: to create a functional and delightful Web environment that is compatible with a sense of personal style	**Goals**: To increase efficiency and productivity of using the system
Strategy: individualisation	**Strategy**: Utilisation
Means: building a delightful Web environment and immersive Web experience	**Means**: Designing, enabling, and utilising useful, usable, user-friendly tools
User model: cognitive, affective, and social, cultural aspects of the user	**User model**: Situated needs of the user
Relational	**Commercial**
Motive: To fulfil a human being's needs for socialisation and a sense of belonging	**Motive**: To fulfil a human's beings needs for material and psychic welfare
Goals: To create a common, convenient platform for social interaction that is compatible with the individual's desired level of privacy	**Goals**: To increase sales and to enhance customer loyalty
Strategy: Mediation	**Strategy**: Segmentation
Means: Building social interactions and interpersonal relationships	**Means**: Differentiating product, service, and information
User model: Social context and relational aspects of the user	**User model**: User preference or demographic profiling; user online behaviour and user purchasing history

Researchers, Fan and Poole (2006), call architectural personalisation building a digital environment, the aim of which is to create a user-friendly sphere by building unique experiences. It consists in designing digital artefacts to best suit the tastes and needs of the user. Architectural personalisation relates mainly to the interface design and the theme is to satisfy the needs, and the basic strategy is to individualise the user experience with the service. Individualisation and meeting the specific needs of the user are key, in contrast to commercial personalisation, which aims to maximise profit. Architectural personalisation in the digital environment aims to find a harmony between the aesthetics individual needs and the functionality of the system / service. Such systems use user models that map the affective, cognitive, and socio-cultural profiles of users (Bedingfield, 2003; Shneiderman, 1987). Instrumental personalisation aims to increase efficiency and personal productivity by creating and delivering the most comfortable and useful tools possible in a way that suits the needs of the user. Contrary to architectural personalisation where function and form balance, instrumental personalisation focuses on the functionalities and usability of the system and treats aesthetics as a secondary issue. An example of such personalisation is, for example, an interactive calendar, which sends reminders about important matters and is able to perform some of the user's tasks on its own (Maes et al., 1999). The implementation of instrumental personalisation is favored by multi-channel and ubiquitous presence - which is an inherent attribute of mobile applications (Nasiri et al., 2018). Another example of instrumental personalisation is personal travel guides (Cheverst et al., 2002; Clarke III, 2008; Sadeh, 2003) or a virtual home environment (Asensio et al., 2019; Tomarchio et al., 2002). Personalisation systems designed in an instrumental perspective are based on contextual personalisation, collected data from all users, such as time, location and environmental parameters, and based on data, and create models that predict user behaviour and make recommendations accordingly. Relational personalisation is based on creating a unique network of social relationships. This approach is related to sociology, communication and anthropology and is based on the assumption that social relationships give individuals a sense of well-being by creating sense of belonging. Relational personalisation is an element that enriches social media platforms. Its main purpose is to help create such a network taking into account different levels of privacy needs. Facebook is an example of such personalisation. The systems filtering the content and recommending friends based on demographic characteristics and place of residence and the system recommending potential friends based on the personality profile (Ning et al., 2019).

The fourth type of personalisation is commercial perspectives - it involves the diversification of the product, service and information systems in order to increase sales and customer loyalty. So far, it has been done through customer segmentation and designating a group of users with similar needs. This type of personalisation is closely related to information technology that enables mass personalisation. The paradigm is shifting from mass goods to highly individualised goods 'building one customer relationship at a time' (Peppers and Rogers, 1993). Commercial personalisation is based on the assumption that the product or service become the more valuable the more they give the user the greater satisfaction (better user experience). The goal for the company is to maximise profit and minimise customer retention. (Riecken, 2000). Customers eagerly use personalised products, personalised services and services with an improved user experience (Ives and Piccoli, 2003). Businesses creates special systems for collecting user knowledge (BI) and relationship management (CRM). The better the knowledge and understanding of customers' motives, interests, behaviours and demographic, sociological and cultural factors, the greater the chance of success of a product or service (Kumar and Desai, 2016; Larsen and Tutterow, 1999).

From the point of view of dissertation goals, neither the personalisation history itself nor the evolution of the concept of personalisation matter. In addition, it seems that this story is still in the making. The definitions are in the crystallisation phase, and the whole process is closely related to technological development. Thanks to digitisation, services have better access to user data (more personal electronic devices). All publications found in the literature review phase of the study focused on specific techniques and personalisation methods and focused mainly on their feasibility and accuracy. Standard personalisation techniques include rule-based filtering, team-based filtering and content-based filtering, web usage exploration. They all have one goal. They use available data and data analysis to understand customer behaviour and deliver personalised products, advertisements or content. All these techniques are linked by the fact that there is data available from a given service.

Concerning personalisation strategies, two are commonly used to accommodate user needs and foster volitional engagement - System Initiated Personalisation (SIP) as a traditional personalisation method and User-Initiated Personalisation (UIP) ((Sundar and Marathe, 2010). In SIP, personalisation is the result of the analysis of automatically collected information from users. Importantly, some of the data is collected in an open manner collected through forms for the user. But at the same time, the system also collects data covertly. But the system stores and processes data related to the user's online activity. By storing cookies, theoretically, in order to adapt websites and their content to the preferences and requirements of the user (Sundar and Marathe, 2010), data from multiple user sessions is collected and aggregated. By this, UIP refers to the empowerment of an individual using an information customisation device based on the identification of user characteristics to increase user's satisfaction and loyalty

(Kalyanaraman and Sundar, 2006; Kreuter and Wray, 2003). Ultimately, customised devices give customers the freedom to adjust the content and features of a particular device directly to their personal preferences, thereby providing them with an active role (Boyle et al., 2008).

Based on the literature review and expert knowledge, three categories of targeted personalisation can be distinguished. They are dependable on the types of data collected about users to create personalised experiences:

- **demographic** - demographic targeting is about finding out who your users are (socio-demographic characteristic). Based on belonging to a group defined by demographic characteristics such as gender, age, education, place of residence, etc., the decision about display or relevant content to the user is taken, and user experience is personalised. Demographic data is collected in many ways, most often by asking the user during app on-boarding process. The knowledge about users demography can be obtained by direct questions form, integration with another platform or taken from CRM.
- **contextual** - contextual targeting is about finding out information about the circumstances of the experience (what device the user is using, the time of the day, or the user's current geographic location) to personalise the app experience.
- **behavioural** - behavioural targeting is about utilising the user's actual behaviour to personalise their app experience. This targeting is executed in almost real-time because it is based on how users behave.

In the article (Fan and Poole, 2006), researchers cited 22 different definitions of personalisation. They are formulated from many different disciplines: marketing & e-commerce, cognitive science, social science, computer & information science and environmental psychology. The most often definitions of personalisation mentioned in this article include the purpose of personalisation and the element being personalised (interface, content, etc.). Also, many definitions have not fulfilled the requirement for scientific definition because they are not neutral (they are favourable to one specific approach). In conclusion, the authors propose an adaption of Blom's general definition of personalisation: *"personalisation as a process that changes the functionality, interface, information access and content, or distinctiveness of a system to increase its relevance to an individual or a category of individuals."* (Blom, 2000)

However, the mere difference in the definitions of personalisation and customisation is of little consequence in determining the purposefulness of this dissertation. The disadvantages of both of these methods are essential. The Table 2.17 compares the advantages and limitations of both methods. The critical problem of personalisation is the time it takes to learn user preferences. Time is proportional to the intensity and frequency of use. The user himself carries out customisation, so it can be implemented from the first moment of use, as long as the user takes the time to define his preferences in the service. However, for services consisting of hundreds of functionalities, such a customisation process is not possible. Moreover, no one can define these preferences with unknown technologies that they have never used before. These techniques can be effective for specifying interests or content but are not definable for more complex advanced services.

Considering the advantages and disadvantages of both solutions, it is clear that neither method can meet the business needs. In personalisation, the problem is the inability to profile the service automatically based on data from the moment of installation, and in customisation, the user's declarative data limits the scope of the profile to what s/he already knows. They do not seem to meet the user's needs either. History-based personalisation may be out of date or not attractive for novelty seekers (high Openness) or inadequate if performed for groups defined by demographics. Within a homogeneous group in terms of gender and age, there is still variation in personality dimensions.

Furthermore, the needs of extreme distribution groups for traits (over 1 SD from the average) differ significantly (Section 2.1.4). In turn, customisation is associated with effort, so it is a cost for the user, and (s)he may become a victim of his/her limitations. To summarise, there is a niche for a method that will overcome these disadvantages while maintaining the benefits. This method will be presented in Chapter 2.5.

2.5.1 Customisation/Personalisation as a Result of Enterprise Strategy

In the previous section, the difference between customisation and personalisation was discussed from the customer perspective, involvement in the service individualisation process and criteria used for classes definition. Nevertheless, in general, customisation, defined as 'fulfilling the customer needs', is the vital element of contemporary manage-

Table 2.17. Comparison of Advantages and Limitations for personalisation and customisation

Method	Advantages	Limitations
personalisation	- does not require any extra effort on the part of the user (the system automatically adjusts content according to user needs) - seamlessly provide users with content that matches their interests; helps filter out irrelevant content, and deliver a better user experience than an unpersonalised service	- the system is learning the user and this process is dependable on the frequency and intensity of service usage - the user privacy issue, some users find 'content tracking' unnerving, and feel uncomfortable with the idea of sharing the personal information - hard to achieve a personalisation system that really successfully delivers relevant content to users (mark missing problem)
customisation	- enhance user experiences by letting user decide exactly what they want to see - users who are afraid of personalisation in the privacy context prefer customisation, because it keeps them firmly in control of their interaction with the service - do not require time for 'learning the user'	- requires a lot of effort on the part of the user, so it has a much higher transaction cost than personalisation – users have to put in the time to set their preferences, personalise their experience, and filter through irrelevant content themselves. - becomes problematic if users do not actually know what they want or they do not even know the possibilities and functionalities of the service - service experience will depend on the technical, mental, cognitive and creative abilities of the user

(Source: own work)

ment strategies. Therefore, before presenting the proposed solution, the role of service customisation from this perspective needs to be discussed.

From the second half of the twentieth century, mass production began to disappear gradually. This change resulted in the necessity to adapt the concepts, principles, and enterprise management methods to the constantly changing environment. One of the main reasons for this increasing volatility was the rapidly evolving technologies. In the contemporary business environment and circumstances, the concepts based on copying proven concepts and management methods, which characterise lean enterprises, no longer work. The concepts of agile enterprises were created that generate value in cooperation with the client, thanks to which they can quickly adapt to the changes. Thanks to the possibility of introducing quick modifications and quickly received information about changes in demand, agile companies can quickly respond to customers' needs. In the diagram 2.3, copied (and translated) from (Trzcieliński, 2020) the key features are shown, and the differences between the four main paradigms of company management are highlighted.

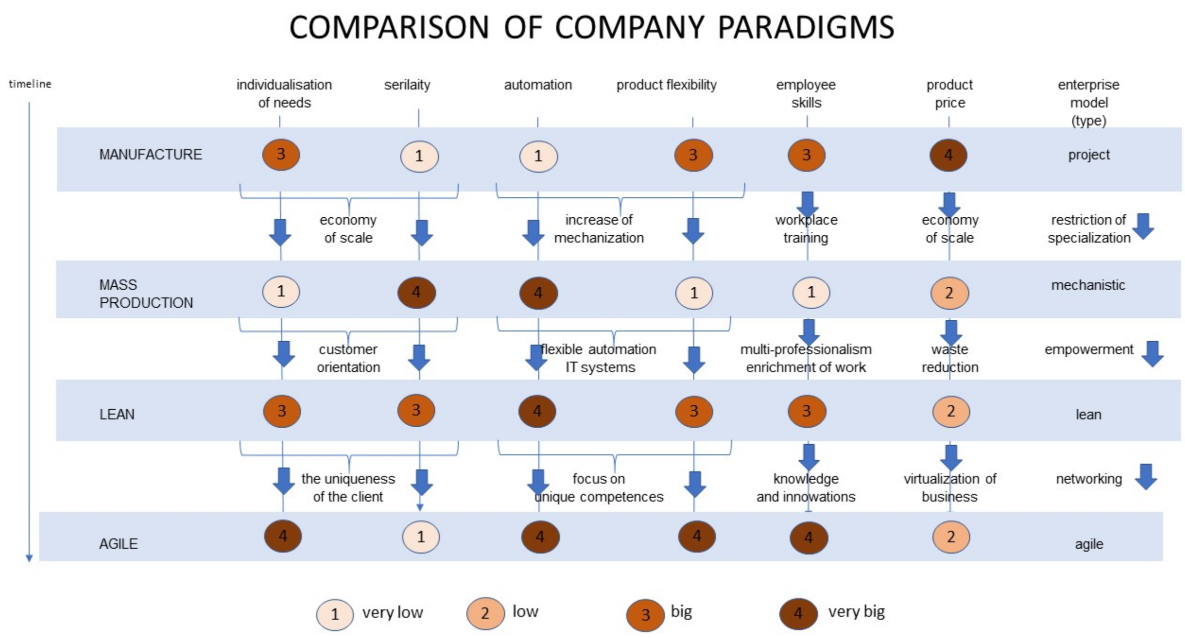


Figure 2.3. Comparison of enterprise paradigm.

Adapted and translated based on (Trzcieliński, 2020), page 14

The diagram shows the general evolution of changes in most important dimensions like individualisation of needs, automation, staff skills, costs, prices. What distinguishes the agile company from the previous paradigms is, above all, a very high level of customisation of needs, low demand for serial production, very high flexibility. In addition, the client's uniqueness, concentration in the enterprise on unique competencies, having innovation and knowledge of modern technologies, and virtualisation and digitisation of the business itself are the phenomena that have set up agile enterprises. Since there is a need to adapt to the increasingly unique needs of customers, the role and the need to customise or personalise and individualise products and services is growing. In the literature on the subject, agility is often given as an element of a lean enterprise, emphasising that its main feature is the ability to quickly respond to customer expectations and redefine or reconfigure value (Trzcieliński, 2020). Prof. Trzcieliński distinguishes four main stages of enterprise development with different paradigms and strategies. The first was the industrial revolution, initiated in the second half of the 18th century, which resulted in the mass creation of manufactures. The second phase was mass production based on the economies of scale achieved through production specialisation. The third is the lean manufacturing paradigm. In this paradigm, the knowledge, intelligence and innovation of employees play a significant role. In this third paradigm, there was a significant reevaluation of man's role in production systems, but the reevaluation concerned only the staff in the production plant. The main reason for changing the paradigms of an industrial enterprise was the constant pursuit of competitiveness of companies in the conditions of a changing environment, i.e. one that permanently changed trends. These changes were caused, among others, by differentiation of consumer expectations caused by both the increase in their wealth and the saturation of demand for mass-produced standard products. Peter Drucker calls lean production 'true mass production' (Drucker, 2012), that is, one that focuses on the principle of mass customisation. A mass customisation is an approach to running a business that increases the adaptation of products to customers' individual needs while maintaining production costs similar to mass production. According to Kurzweil Law of Accelerating Returns, technological development takes place at an exponential rate (Kurzweil, 2004).

As a result of this technological acceleration, they increase the variability and unpredictability of the environment. Therefore, it is reasonable to ask what enterprise model will recognise such an environment as favourable, not hostile. In lean companies, the following elements of organisation processes can be distinguished: product design, technology design, supply, production, sales. Furthermore, as a lean philosophy, it distinguishes the following principles: define value, identify value stream, flow, suction, improvement. There are also three basic concepts to help business: reduce costs, improve quality, and shorten customer needs. On this basis, the concept of Total Quality Management (TQM) was created. The concept is to organise an enterprise to manufacture products or provide services that meet customer expectations. Satisfying customer needs becomes the primary goal of the company's operations (Armstrong, 2007). Such companies focus on the customer (organising user-centric). In turn, in the first years of the twentieth century, the concept of an agile enterprise developed. The name is closely related to the feature that is expected from the enterprise (agility) and which it aims to achieve and thanks to which it will *'gain the ability to immediately respond to changes in the market environment by using the key resource, which is knowledge'* Kidd (1994, p.29-30). 'Opportunity' and the ability to seize of an agile enterprise. They are primarily the needs reported by customers and created by the company.

2.5.2 Delayed Individualisation as form of Postponement Strategy

In the previous chapter, the basic features of a modern enterprise were described, emphasising the role of adjusting the final products to customers' needs. In combination with the environment high volatility and demand variability, this resulted in the need to develop a strategy to counteract errors resulting from the too long time between creating the product concept and the moment of its delivery to the Customer. The literature on the subject describes various logistic strategies of enterprises and supply chains, and one of them, commonly applied, is *postponement*. The main reason for using the delay strategy is to gain a competitive advantage by delivering the most up-to-date personalised product to the Customer. The principle of postponement has also been termed as late customisation or delayed product differentiation' (Swaminathan and Lee, 2003), causing the consumer to be more satisfied with the product. However, at that time, the concept of postponement concerned only the moment of assembling the final product in the factory. Subsequent phases of the concept's evolution up to the beginning of the 21st century brought many new concepts of deferral in the supply chain. Professor Hoek stressed the need for much new research in the area of postponement tactics used. According to him, the deferral related to production and logistics should be investigated and the deferral in the context of delivering and customising services (Hoek et al., 1999; Van Hoek, 2001). According to Swaminathan and Lee (2003), there are three types of factors that affect the benefits and costs of deferral - market, process, and product factors. Market factors are related to the customer demand and service requirements. Product and process are less important from the point of view of this dissertation. The effectiveness of a deferral strategy depends on how well the company can adapt its process and product characteristics to the requirements of the market and the consumer (Swaminathan and Lee, 2003).

Postponement is a strategy prepared for management in a situation of uncertainty.

'Market uncertainty may result from ignorance of market parameters (e.g. the price of a good, parameters of the demand or supply function), or may relate to the behaviour of competitors and partners in market activities' Stefan (1996, p.38). Uncertainty is also directly correlated with risk. The higher the uncertainty, the higher the risk, and the risk becomes one of the key factors influencing business activity.

Quoting for (red) (2012), (translated from Polish), *'Management strategies, adopted in conditions of uncertainty, most often focus on two ways of proceeding:*

- *1) Seeking opportunities to increase certainty, which means focusing on minimising uncertainty.*
- *2) building uncertainty into management and creating a flexible, self-regulatory system based on the principle of operation of fuzzy sets or neural networks.'*

Today's technologies already enable the development of such a flexible, self-regulatory system, based on neural networks and using the available data to create knowledge at the client's place, will be the subject of the present dissertation. Such a system will require a standardisation of the process of preparing such a self-regulatory application. However, it falls within the scope of postponing the creation of the final shape of the product until the service (application) is installed at the Customer's site.

2.5.3 Data Flow as a Result of Personalisation Enterprise Strategy

The previous chapter described and compared various data-driven personalisation methods. In this section, there is a brief discussion about the relationship between the way and type of personalisation and the logic of the data used for this purpose. With the development of technology, personalisation is becoming an increasingly important segment of the economy. In an enterprise, personalisation complements customer relationship management (CRM). CRM is seen as a strategy for attracting, developing and retaining customers. Personalisation is an approach that can help attract the customer (personalise ads), keep the customer (personalised customer experience), and bring the customers back to the service.

The overarching goal of the personalisation strategy is to deliver the correct information, products, services and data to the end user (customer) at the right time and place. The aim is to provide a competitive advantage to the market. Data sources and

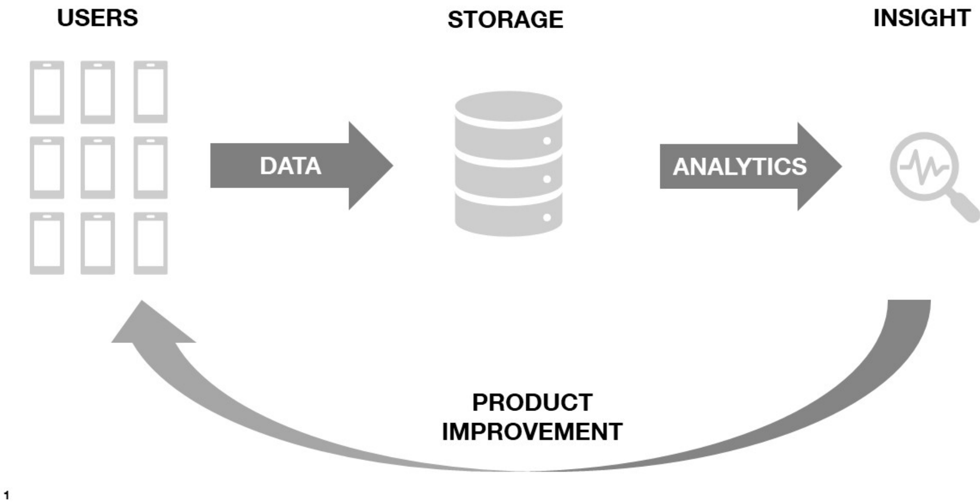


Figure 2.4. Diagram of Data Flow in Traditional Business Intelligence

(Source: own work)

data processing flow determine the choice of personalisation for a given service or product. In the traditional Business Intelligence approach, data from the use of the service is stored in the Enterprise Information Systems, Data Warehouses, and CRM Systems.

The diagram Fig. 2.4 shows a traditional enterprise data flow that aims to use customer knowledge to design and improve products and services. In this traditional approach, knowledge about the customer is separated from the customer's end device. Data is collected, analysed, and decisions are made in a more or less automated way based on the analysis of the collected data. An example of personalisation in this traditional data flow model would be to extend the service with new functionalities based on satisfaction surveys for specific groups of users recognised on the basis of data available in CRM, such as gender, age, place of residence or the portfolio of used products. Or personalised recommendations for the displayed promotional materials (targeted campaigns). In this data flow, the process of obtaining customer information and analytic is completely separate from the decision-making system initiating the change. Most often, changes to the offer are initiated by a human. According to Božič and Dimovski (2019), BI remains the main priority in the information strategy in order to increase sales and operating results.

The next rutting diagram Fig. 2.5 shows a different data flow in the case of personal-

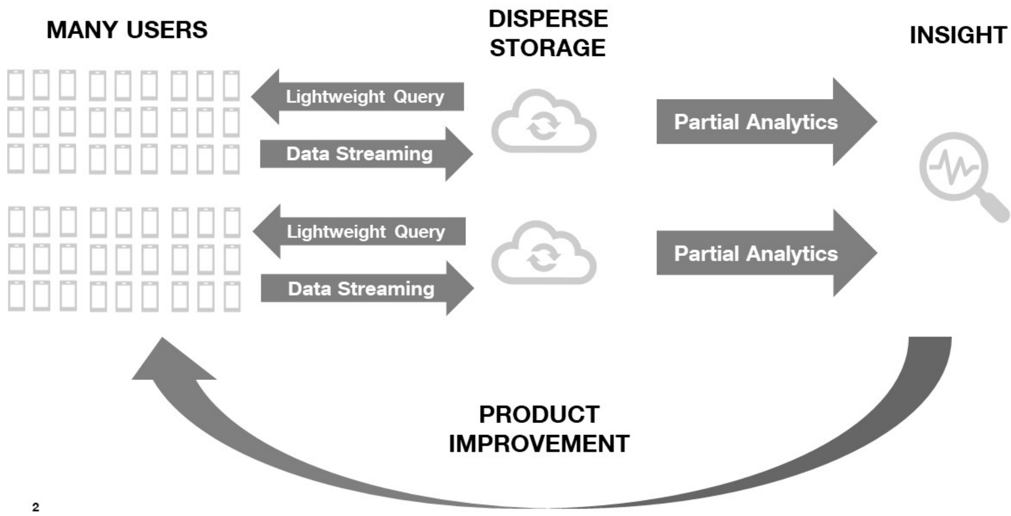


Figure 2.5. Diagram of Data Flow in Big Data Business Intelligence Approach

(Source: own work)

ising electronic services based on Big Data analytic and the use of distributed computing in the data cloud. Thanks to the availability of technology that enables the processing of data from various sources and easy communication between services and information systems (API REST technology), it becomes possible to create systems that make decisions based on data almost in real time. Thanks to such a system in which data is sent to various analytical components, contextual, individualised and targeted personalisation is possible in terms of place, time and person. Such complex information systems are already fully automated and even to a limited extent autonomous. What is important for traditional data processing and multi-source data processing, characteristic for Big Data systems and Intelligent Information Systems - the data is duplicated and saved in many places. A user profile, no matter how defined, is permanently available and usable by the enterprise.

2.5.4 Personalisation Data Privacy Paradoxes

Automated personalisation of services based on digital data inherently requires processing and the collection of personal data of service users. Despite attempts to regulate data protection from the legislative side (GDPR provisions), there are many doubts about their effectiveness due to their limitations (Finck, 2021).

There are two separate paradoxes related to data protection: the first concerns the inverse relationship between minimising access to data and the quality of algorithms created on this basis, and the second is related to the fact that the protection of data used to create customer knowledge does not fully protect user data created based on their basis. First, there is, therefore, a conflict between privacy and personalisation; the more information the system has about the user, the more effectively it can adapt to his needs, but as a consequence, the privacy of the user's data is less protected. This situation has been called the personalisation-privacy paradox, and many studies are addressing this problem, although usually in the context of web services and content recommendation systems (Awad and Krishnan (2006), Egger and Rainoldi (2021), and Idberg et al. (2021)).

Publicising data leak scandals makes users more aware and prudent. In addition, legal regulations requiring the collection of consents from users mean that users know when services use their data . On the other hand, the repeated and tedious mechanism of frequent consent granting causes that consent to use data is given without reflection and reflexive action. Regardless of which type of user we are dealing with, whether the more cautious and distrustful or on the other hand the reckless, consenting user, personalisation is crucial in industry 4.0, and trust is an essential element of its development. It seems that a solution enabling the personalisation of the service without the need to record and send user data beyond the end device may be an effective antidote to the current problems.

2.6 Shortcomings and Gaps in the Existing Research

The review of the literature on methods of determining personality from data and methods of automatic personalisation, presented in this chapter, revealed gaps in the existing solutions. To synthesise the review results, the presentation of main shortcomings concerning existing works will be defined in the following points.

1. Determining the user's personality based on data is in most cases based on data collected when using a specific service (this mainly applies to social media). In most published cases, the model is built from social media data. The same method of collecting the data on which the profile is calculated takes time to record the

behaviours relevant to the model. Neither of these methods can calculate the user's personality profile at the time of service installation from the available data. Therefore, their usefulness and relevance is not related to the method of calculating the personality profile from the data, but to the availability of data to calculate this profile in order to adapt to the user's needs as early as possible;

2. Determining personality in any of the presented ways does not ensure complete privacy of the user, i.e. it is counted on personal data, sent to external servers, repeatedly recorded outside the end device. Therefore, by using the service, the user loses his rights to privacy protection. Data storage and security have not been raised in the literature on the methods of determining the user's personality based on digital traces. The authors of only one data-driven personality study ((Bin Tareaf et al., 2019)) from a profile picture wrote openly that their goal was to raise awareness among social media platforms about using their private data by third parties.
3. Another research gap identified in the literature review concerns the very method of personalised services. The analysis shows that there is as yet no method for flexible automatic customisation/personalisation of the service. What is commonly practised is the use of data from service usage to improve the user experience. However, there are no automatic methods of personalising the user interface or communicating with it from the moment of installation. The literature review shows that no method combines the benefits (for the user) of automatic profiling and automatic personalisation of the service.
4. Moreover, there are no studies that would constitute an adequate reference point for the studies and methods proposed in this dissertation. Both in terms of user data (data from the mobile phone) and the way it is used (counting the model during the installation of the service from the data accessed by the user, without transferring this data beyond the user's end device). Only one publication reporting research on the model based on data from a smartphone was identified, but the data was collected for 30 days of observation, and the model was not based on statistics but events logs (Stachl et al., 2020)
5. In the literature to date, little attention has been paid to discussing ethics and rules for using user data in the service. Privacy is undoubtedly a subject that requires

regulations and studies also from the legal point of view.

6. Little attention has also been paid to personalisation based on the personality profile and its use in activities that benefit the user, not just in tailoring marketing communication. This issue seems particularly important in this dissertation. The development of technology enables much more reliable and people-friendly solutions.
7. Another observed gap is inappropriate and inconsistent with psychological theory, is the purely statistical approach to creating classes based on personality questionnaires. Most of the research is based on binary labels, which facilitates model building but disqualifies the use for personalising the service.
8. Finally, it seems that the current personalisation methods do not meet the goals set by the postponement strategy characteristic of modern agile businesses. The customer receives a service targeted at a typical user, and the existing personalisation methods adjust the service while using it. However, this strategy has no chance of success if the service is rejected in the 'typical' default version. Developing personalisation methods from the first contact with the service would solve this gap.

2.7 Summary of Literature Review

This chapter presents the methods of personalisation and the theoretical foundations of personality theory, and their implications. It has been shown that calculating the personality profile, due to the access to a large amount of digital data, is possible and valuable for inferring about users' needs (resulting from behavioural motivation and deficiencies and dysfunctions hindering functioning). So far, data-driven personalisation has been relatively widely used despite the disadvantages identified. The environmental analysis was carried out both from the point of view of business analysis (no solutions enabling automatic profiling from the moment of service installation) and from the point of users of advanced services and mobile applications existing on the market. personalisation seems to be used from a business point of view, but due to the methods used, it is not always perceived as beneficial for users (matching to the history of use is not always adequate, problems with the use of sensitive data). The user's personality seems valuable information about the user's needs, although it must consider the actual differences between individuals (emphasising the non-specificity of needs) and no statistical differences between group means (binary classes). Unfortunately, there are no solutions that combine benefits for the user to protect his/her private data. Additionally, personalisation seems to be a key aspect in building the competitive advantage of a modern agile enterprise. From this point of view, the search for automating both the process of obtaining information about the user's needs (determining personality based on data) and adjusting the service to these identified needs seems to be an attractive area for exploration research. From this point of view, the search for automating both the process of obtaining information about the user's needs (determining the personality based on data) and adjusting the service to these identified needs seems to be an attractive area for exploration research. This research can enable the complete application of the postponement strategy. The final product will obtain the final form for the user when the service is provided, i.e. during the installation of an electronic service, e.g. a smartphone application.

2.8 Concept of Personality Aware Services

Chapter 2.3 presents the two main approaches to customising services to user needs: personalisation and customisation of services based on digital data, and discusses the advantages and disadvantages of both approaches. The key aim of this dissertation is to propose a different approach that combines the advantages of both presented methods and eliminates the disadvantages at the same time. Furthermore, as customisation does not take place automatically based on user data, the essence of the new approach will be presented by comparing the course of the process currently used in the automatic personalisation of services (Fig 2.6) and the proposed approach (Fig 2.7). Finally, the new approach (Fig 2.7) will be compared with the service customisation process done directly by the service user. The most common automated data-driven personalisation model is currently creating a user profile from the collected data (fig 2.6). It is done either by tracking the user behaviour based on the user's tags (websites) or by collecting data on user activity - when profiling is done for one service (e.g. social networks, TV platforms, sales applications, banking applications, etc.). This profile is adaptive,

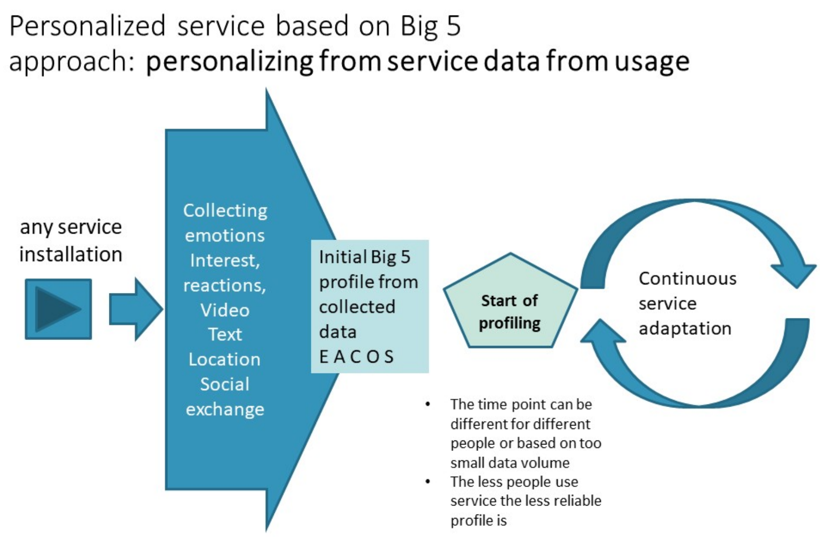


Figure 2.6. Diagram of Old Idea of Service Personalisation

(Source: own work)

depending on the length and intensity of service use. Using a profile for an individual user and offering them a personalised experience with the service may take place at different points in the product life cycle. Moreover, by design, it is rarely available from the moment of installation. The exception may be personalisation based on the declared or actual demographic description. In this process, the Big 5 profile could only be counted after some time of use. It is also important that the less the user uses a given service, the less probability of calculating the correct profile is.

The main distinguishing feature of the new personalisation proposal for any interactive service is creating a user profile during the service installation process. Moreover, each of the personalisation methods discussed in Section 2.3 involves the processing and preserving a large amount of sensitive data about the user. personalisation can be based on various data types related to consumption, personal activity, location, emotion detection, interest and content tagging, social network and many other data (Section 2.2.1).

The proposed new personalisation method is based on statistical data related to using a smartphone found on the device at the time of service installation. During installation, based on the statistics from the data detailed in Chapter 4, a user profile is created describing the user's Big Five profile on a normalised sten scale. Profile availability from

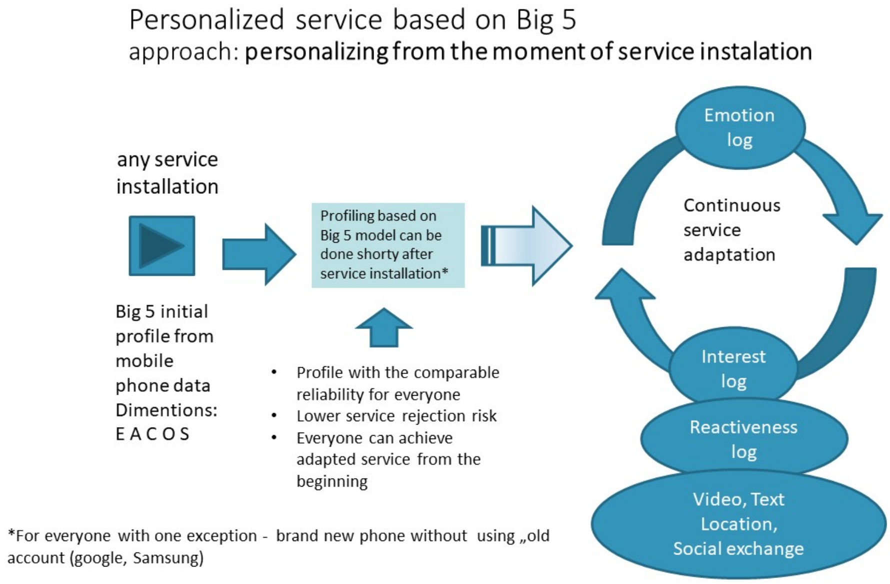


Figure 2.7. Diagram of New Idea of Service Personalisation

(Source: own work)

the moment of installation gives the possibility of service adaptation to the psychological and personal needs of the user. The profile remains part of the front-end of the service and remains private, only available to the user version of the site. As a result, Introverts will receive a service that looks different from Extroverts. Highly Openers will receive a different recommendation algorithm, taking into account their exploration needs. Moreover, the lowly conscientious will receive a service that will compensate for their voluntary nature and reluctance to plan. Similarly, a voice assistant or chat-bot communicates in a manner tailored to the needs of each group. That is why this category of services can be named a service aware of the user's personality because it can benefit from personalisation concerning the user's personality from configuring the personalised process of service adaptation and the exposure of the personalised interface. As a result, users can receive this service in a different form, appearance, content and configuration depending on the needs assigned to the personality profile. Having this initial profile does not preclude further use of standard personalisation methods. The initial profile should be verified by assessing the user's choices with the initial profile. Data from the service users will be used to auto-calibrate the profile, thus increasing its fit. Chapter 4 will discuss creating such a personal profile based on data from the telephone and personalisation based on it. The main distinguishing feature of the new personalisation proposal for any interactive service is creating a user profile during the service installation process. Moreover, each of the personalisation methods discussed in Section 2.3 involves the processing and preserving a large amount of sensitive data about the user. personalisation can be based on various data types related to consumption, personal activity, location, emotion detection, interest and content tagging, social network and many other data (Section 2.2.1).

The proposed new personalisation method is based on statistical data related to using a smartphone found on the device at the time of service installation. During installation, based on the statistics from the data detailed in Chapter 4, a user profile is created describing the user's Big Five profile on a normalised sten scale. Profile availability from the moment of installation gives the possibility of service adaptation to the psychological and personal needs of the user. The profile remains part of the front-end of the service and remains private, only available to the user version of the site. As a result, Introverts will receive a service that looks different from Extroverts. Highly Openers will receive a different recommendation algorithm, taking into account their exploration needs. Moreover, the lowly conscientious will receive a service that will compensate for their voluntary nature and reluctance to plan. Similarly, a voice assistant or chat-bot communicates in a manner tailored to the needs of each group. That is why this category of services can be named a service aware of the user's personality because it can benefit from personalisation concerning the user's personality from configuring the personalised process of service adaptation and the exposure of the personalised interface. As a result, users can receive this service in a different form, appearance, content and configuration depending on the needs assigned to the personality profile. Having this initial profile does not preclude further use of standard personalisation methods. The initial profile should be verified by assessing the user's choices with the initial profile. Data from the service users will be used to auto-calibrate the profile, thus increasing its fit. In Chapter 4 process of creating such a personal profile based on data from the telephone and personalisation based on it, is presented in details.

The previous sections discussed the importance of personalisation for a modern enterprise and the postponement strategies for configuring the end product as late as possible. Delaying actions are taken to adapt the product to the most current needs of the client. The proposed personalisation method, in which the product adapts to the client's personality at the time of service installation, is an example of implementing this strategy. The search for the differentiation of needs and the non-typical nature of individual customers seems to be a better method of individualising the final product. Automating the entire process of tunning a product for a specific customer allows for mass delivery of personalised services. The proposed solution does not eliminate the need for application functional and technical development, but it also gives the possibility of a flexible way to adjust the product already used by the customer. Finally, it is crucial to leave the user's data, including the profile itself, in a private sphere that the service provider has no access to without consent. The user for whom data protection is an essential factor not deprived of the benefits of personalising the product.

Chapter 3

Scheme of the Research

3.1 Introduction

After analysing the existing literature related to personality theories, the existing methods of indicating personality based on data, the role of personality in personalising services and methods of personalising electronic services, as well as after identifying the gaps in these areas, the following chapter presents the proposed research process which allows for creating proposed solution - Personality Aware Service. The first artefact defined in this research is to propose an effective and accurate method of recognising each user's personality (Big 5), based on the available data at the time of installation, without waiting to collect data from the user's history. The second is to use this classification for personalisation, for example, of an electronic service. Finally, the third artefact will prove that personality can be effectively used to classify users' needs as a base for personalisation. This chapter will describe the scheme and method of the entire study concerning the design science methodology. Firstly, the planned procedure, methods of collecting data and creating research tools will be presented. Then, the data sources needed to achieve the study's objectives will be discussed, and the tools (questionnaires and the application) will be described. Finally, the data on which the study is based will be explored further in terms of their quality.

3.2 Research Procedure

The primary objective of this doctoral dissertation is to develop a novel method of personalising interactive electronic services like smartphone apps or AI services like chatbots or virtual assistants. The following research steps were designed to achieve the research objectives (Fig 3.1).

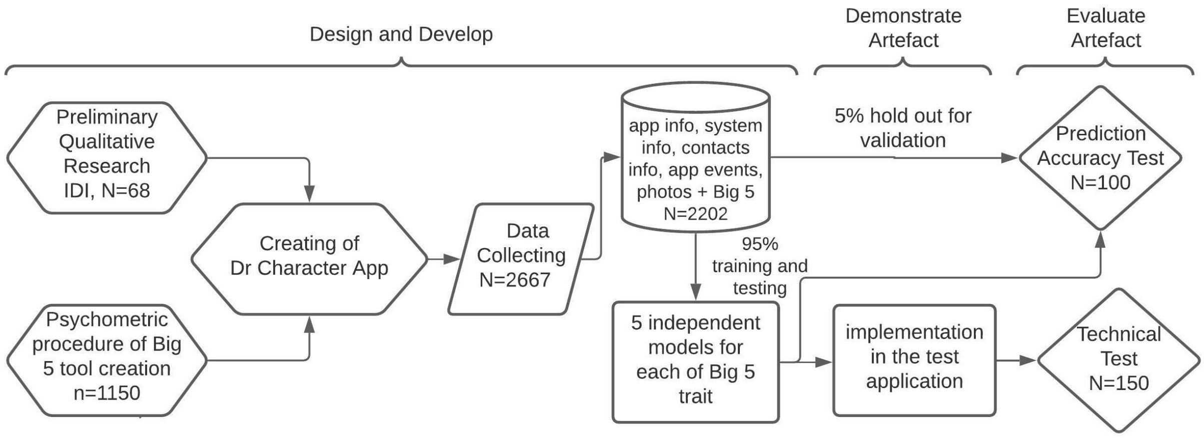


Figure 3.1. Research Scheme & Methods

(Source: own work)

1. **Preliminary qualitative tests** carried out to identify needs and collect descriptions of discriminatory behaviour. The detailed report from this part of the research is presented in Appendix A and discussed in chapter 3.2.1.
2. **Creating the personality assessment tool**, which can be used for this specific research (low number of questions (25) and tool tested and created for online or mobile app usage). Standard psychometric procedures are applied. The detailed report from this part with a standard psychometric report is presented in Appendix B and discussed in chapter 3.3.1.
3. **Developing the mobile application** dedicated to collecting data from mobile phone and Big 5 assessment. So far, the data from the following user activities (sources) have been collected and analysed on a mobile phone of a user:
   - Telco data, e.g.call logs statistics, text messages logs statistics (collected from mobile phone logs)
   - Application data, e.g. applications on mobile phone, phone parameters.

- Photos, e.g. photos in user photo albums.
- Phone settings and statistics, e.g. battery consumption, kind of security level.

The raw data collected is presented in Appendix C and separately discussed in Chapter 3.3.

4. **Main research fieldwork** - data collecting required for creating the model. The short report from data collecting fieldwork with all details is in Appendix D, and methods of data processing, quality check and descriptive statistics are presented in chapter 3.4.
5. **Creating the personality model** based on the mobile phone data, available at the moment of service installation (a single drop without the additional collection of the activity logs). The model creation procedure is the pivotal part of the first artefact creation. The entire cyclical process of reaching the final model is described in detail in Chapter 4 of this dissertation.
6. **Validation stage** at the stage the 2nd and 3rd artefacts are created. That is the interactive service that allows experimental manipulation based on personality adapted service. This stage of the research is described in chapter 5.

The planned comparison of the result based on behavioural metrics and the personality test results can allow for valuable conclusions. It can be said that personality questionnaires present the image of how we see ourselves and not necessarily our real behaviour. If this image is biased (for various reasons, e.g. high need for social approval or cognitive defence mechanisms), the diagnosis will not be accurate. Moreover, the answers in the personality questionnaires are sometimes based on subjective perspective, e.g. self-assessment of the amount of time spent with friends and in social gatherings may not be adequate to the real place in the distribution for the population. So it's possible that during creation the model that predict personality questionnaire results (declarative), this error can be repeated. Before the primary research, requiring large expenditures, a small preliminary study was carried out to verify the validity of the adopted assumptions and directions of the research.

3.3 Preliminary Research

This section presents the key findings from the analyses conducted based on the material collected in preliminary tests. The objectives for the initial study were as follows: (i) to verify personality according to the Big 5 theory as a good classifier for the segmentation of attitudes concerning intelligent services (virtual agent), (ii) to check statistical methods that can be automated in created services for a personalised response. The detailed report and collected qualitative material from this stage are presented in Appendix A.

3.3.1 Method of Data Gathering

The research was carried out on 60 users of mobile phones, residents of Warsaw. In order to emphasise the diversity among people resulting from their personalities, the study was conducted on a very homogeneous group in terms of demographics (the aim was to reduce confounding factors). The age has been limited to 20-29, and groups for both sexes have been separated. The members of both groups were technologically advanced and using many functionalities of mobile devices. The study was multi-stage: first, filling out the personality questionnaire (Big 5), then monthly observation of behaviours concerning social profiles and telephone, and finally in-depth interviews aimed at getting as much information as possible about behaviour patterns. Additionally transcription of these interviews was used as the base for linguistic analysis (Natural Language Processing). The resources for NLP consist of over 11 000 sentences (59% of which were spoken by women) and almost 140 000 words (56% of which were spoken by women). The results of these studies were also used for creating preliminary regression models from available phone data and social media (Facebook) behavioural metrics for each of the Big 5 dimensions. The linguistic analysis exploring differences based on personality are not in scope of this dissertation, but they were published in article (Krzeminska and Rzeznik, 2021).

3.3.2 Differentiation of Attitudes Towards Intelligent Services

A qualitative investigation of respondents' statements confirmed a wide variation in attitudes towards intelligent services among people with different Big 5 profiles. A summary of these declared attitudes is presented in Tab. 3.1. It shows mainly qualitative differentiation of attitudes and expectations of mobile phone users towards virtual agents and services based on AI.

Table 3.1. Different attitudes and expectations connected with virtual assistant (VA) contingent on personality dimensions (Big 5). (Key findings from experiments on mobile phone users - Preliminary Qualitative Research, N=68)

Personality dimension	High level	Low level
Extraversion	rejection of VA, but can be considered if adapted to the user expectations (no advertisements)	rejection based on privacy (need for high protection), VA in the role of an invisible friend
Openness to experience	the repeatability based on usage history is irritating, but the idea of a smart, proactive VA searching for all they need is attractive; they would be first to try it, their interest is changing and evolving	they expect to do something more manageable and faster - but still only when they need it (they initiate the action) - their interests are relatively stable.
Conscientiousness	the only group which likes personalisation, VA for them can be a helpful tool to control themselves, keep schedule and order	they are aware of their problems, so they want the functionalities that help them avoid errors due to lack of systematic or discretionary decision-making
Neuroticism	they appreciate the convenience of this solution (it limits forgetting, helps), but are sceptical about the data transfer and the risk of "manipulation"	if they have enough time for adaptation, they can go with it (the value for them is the effectiveness in real life))
Agreeableness	they are afraid, but if it becomes popular (recommended by friends), they will accept it	their main motivation are pragmatic needs (usefulness in everyday life)

(Source: own work)

The differences concern a virtual assistant's function and are a simple derivative of the diversity of needs. At the same time, it can be seen that attitudes polarise according to personality dimensions. For example, people who are open to experience expect non-standard content, have a high level of cognitive and exploration needs (curiosity). In turn, people with low Openness expect only a sense of comfort in a world that is well known to them (they only like what they know, they are afraid of the unknown). For another dimension, individuals with a high level of Conscientiousness use mainly the functionalities that help implement the need for control, which mainly manifests itself in the control of time and careful planning. On the other hand, people with a low level of Conscientiousness, who accept life in chaos and disorder, need only basic control functionalities and will never be interested in using the advanced calendar or notebook functions.

3.3.3 Summary from Pre-research Part

The conducted qualitative preliminary research provided a wealth of material facilitating the operationalisation of the behaviour of telephone users. A wealth of empirical material was collected, which, despite its non-representative character, has enriched the understanding of the motivations that drive service users. They largely coincided with the theoretical assumptions and the description of the dimensions of the Big 5. It should be remembered that the observed differences were reinforced by a selected target sample (Appendix A), and thus they relate to the differences between the extreme levels of features (deficient vs very high). These differences will not be readily observable for feature levels closer to the centre of the dimensional distribution. The preliminary research suggests that personality can significantly facilitate service profiling as the predictor of user's needs. Based on observing behaviours related to mobile phone usage and experience with different electronic services, the results confirm personality theory in the Big 5 approach. It seems that extending the range of diagnostic tools measuring users' natural needs (connected with the Big 5 profile) can be helpful in the development of intelligent and sensitive services.

3.4 Creating the Personality Assessment Tool

The completed report from the procedure and analysis is available in Appendix B of this dissertation. Therefore, this chapter presents only a discussion on the legitimacy of creating this tool and the psychometric value of the obtained tool. Due to the specificity of the research and linking it with electronic mobile services, creating a dedicated Big 5 tool was undertaken. The available diagnostic tools for personality measurement have a licensed form and are paid (e.g. NEO Personality Inventory-Revised and NEO-FFI

(NEO Five-Factor Inventory), created by Costa and McCrae (1992) and adapted into Polish language separately by Zawadzki et al. (1998) and Strus et al. (2014). These tools are used primarily for individual clinical diagnosis and assessment. Psychological diagnostic tools cannot be disseminated because it weakens their diagnostic power, nor be modified. The procedure of using them is rigorous and can be conducted only by authorised psychologists. Another reason was the need for integration of the Big 5 tool with electronic applications. The traditional tools are created and normalised in paper and pencil circumstances. Thus, for validation purposes, the kind of academic tool was used. It was the Polish adaptation of IPIP-BFM-50 carried out by Strus et al. (2014). However, 50 questions in IPIP-BFM-50 were too numerous to use directly in the main study. With the acceptance of possible worse psychometric parameters of the shorter version, the decision was made to create an independent tool, possibly the shortened version. Creating a new tool was also beneficial for other reasons. First of all, there was a requirement of having the possibility of applying it in the electronic version, which would be integrated with the application. As a result, we have complete data necessary to create other tools for different needs that may contain more or fewer items or are limited to only one or two personality dimensions. In the case of ready-made tools, there is no such possibility. The whole psychometric Big 5 tool creation procedure was processed from the beginning. In this way, the creation of a dedicated tool became possible and easy. Due to business needs, personality dimensions: Extraversion, Agreeableness, Conscientiousness, Neuroticism (Emotional Stability) and Openness to Experience, will be treated independently - as if we were creating five separate tools. The issue of orthogonality of factors will also be examined due to theoretical foundations and the problem of interference between factors in building a model based on data.

3.4.1 Data Gathering for Psychometric Part

The assumption made for the psychometric procedure concerns most of all the aim for which the tool is created. The tool will be tested in the electronic version of the survey (online research), distributed via a website [www.szoprojekt.pl]. The IPIP-BFM-50 survey was also carried out in the electronic version, and its equivalence of measurement in different survey conditions (paper and pencil versus online survey) has been confirmed through Confirmatory Factor Analysis. It should be noted that the final tool will not be used for diagnostic purposes but for research and classification of the trait levels (low, medium or high), as well as to define the dimension dominant for a specific person. The research was carried out on a group of volunteers. The link to website with the survey was distributed via Internet communication, mainly through different sorts of social media, such as Plazza and FB. The research was conducted between February 2019 and May 2019. There were 1333 surveys collected altogether, and 1150 of them were qualified for the research itself. The eliminated surveys were considered 'clicked' without consideration, as the time spent filling them in was too short to allow careful thinking. The participants were to complete two surveys: the first one contained 60 questions – proposals, among which the final items of B5PF items were to be chosen, and the second one contained 50 questions from IPIP-BFM-50 (Strus et al., 2014). The demography was limited to two questions about gender and age (see Appendix B for details).

3.4.2 Summary from Psychometric Part

Finally, two versions of Big 5 tools were created B5PF-5 and B5PF-6. The first, shorter one named B5PF-5consists of 25 questions in total (5 questions per Big 5 dimension), and the second, longer one (B5PF-6), includes 30 questions in total, (6 questions per Big 5 dimension). From the conducted analyses and statistics (see for details Appendix B) it can be concluded that both versions of the questionnaire for the study of 5 great personality factors (Big 5) meet the essential criteria for psychometric scales for scientific purposes. Satisfactory parameters were achieved for 4 out of 5 factors. The weakest though still acceptable, is Agreeableness. Both 25 and 30 question versions can be used to classify and rank dimensions to determine clients' dominant personality profile features. In case of having more empirical material, all the above analyses can be repeated to determine the current psychometric parameters of the tool. Repetition of analyses is particularly recommended for normalising results. Undoubtedly, further examination (on larger samples and in other studies) is whether the parameters for the Agreeableness coefficient are not a consequence of the high skewness of the distribution of natural results (high social approval for conciliatory behavior significantly distorts the results, shortening the scale length). For further leading research, the shorter version (B5PF-5) was used.

3.5 Data for Creating Personality Model

In this subsection, the method of collecting data for research will be presented, the tools used for this purpose, as well as the data themselves and their properties will be discussed. Due to the defined research need, to achieve the defined goals and the first artefact, data on which the personality model will be based are necessary. For this purpose, a phone application called *Dr Charakter* was created, and it was used to collect anonymous data for research.

3.5.1 Application for Data Gathering for Creating Model

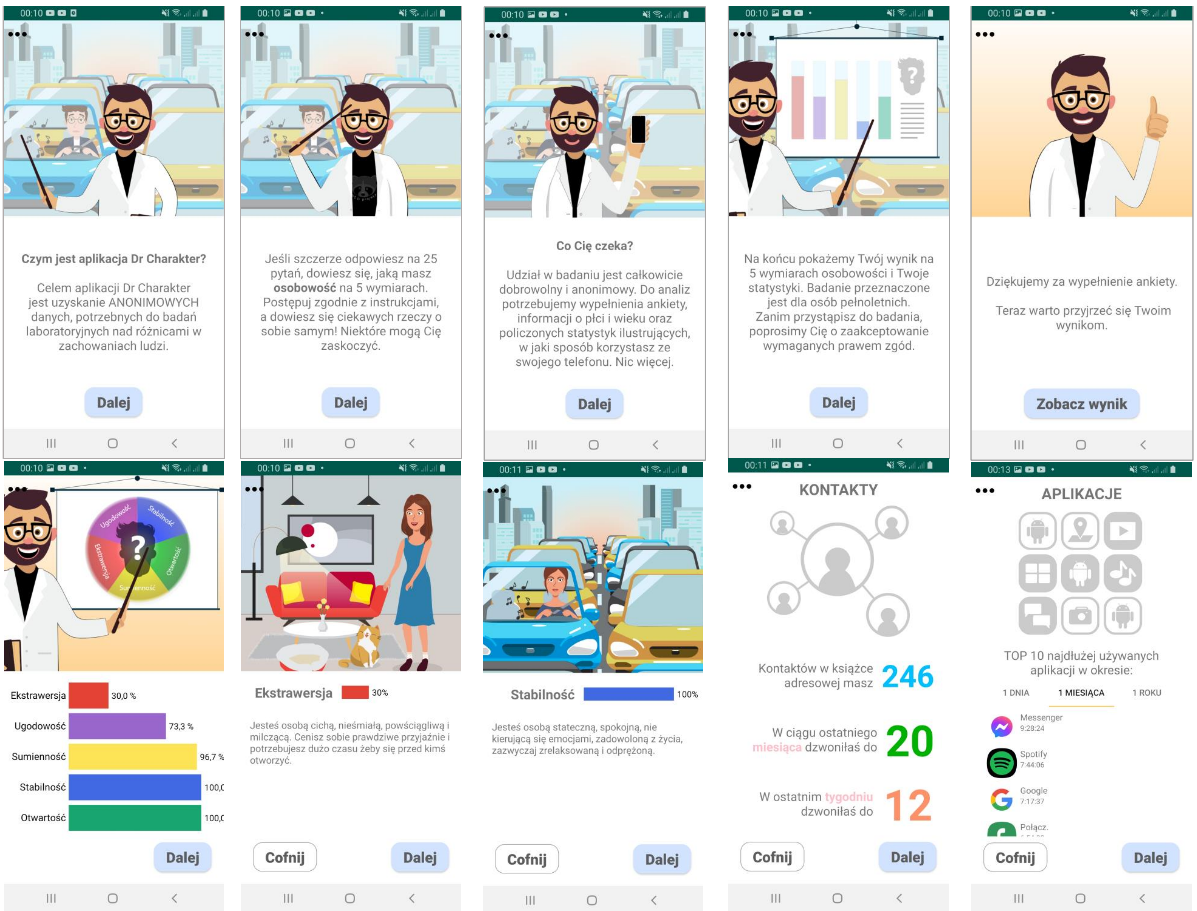


Figure 3.2. *Dr Charakter* - Example screens form App for Data Gathering (own source)

It is an application running on Android, and thus the test can only be carried out on phones with the Android operating system. Application was named *Dr Charakter* and picture 3.2 is presenting a few sample screens from the application.

Figure 3.2 shows a set of Dr Charakter screens from the course of the study (In

Polish). The study procedure looks as follows:

1. Explaining to the respondent how the application works.
2. Messages and explanations accompanying granting the necessary consents for access to data in the telephone
3. After granting the approvals required by law, the application starts counting anonymous statistics.
4. The respondent answers 25 questions from the personality questionnaire. (5 questions per dimension). The questions are formed based on behaviour description. The respondent's task is to estimate on the 7-point LIKERT scale how closely this description fits his/her behaviour. One means 'definitely does not fit' and 7 'definitely does fit'. Thus, means of the scale is symmetrical with the neutral centre in point 4.
5. Then the respondent declares his gender and age.
6. The person receives the personality questionnaire results in the form of the personality profile of the subject in the form of a bar graph on a 1-100 scale, intuitively understood as the% intensity level of a given trait. The presented results are raw, without normalisation to the population.
7. The respondent answers seven additional explanatory questions.
8. At the end, basic sample statistics calculated from the data from the phone are presented.

3.5.2 Description of the Data Used for Creating Model

Finally, filled Big 5 questionnaires and smartphone data from 2666 people were collected. In total it was 17,3GB of data, including descriptive statistics about: photos (12GB), exifs (1,4GB) and call/ text messages/ application/ smartphone data (4,6GB). The complete list of data categories is presented in Appendix D. The Table 3.2 defines data categories and contains examples for all 9 data categories. There were over 200 of the raw data types. Standard Android Information data category is most numerous, and available for every installed app without dedicated consents.

All other data require special and dedicated user consent to be available for the app. The consents needed are:

Table 3.2. Descriptions of the Data Categories Taken from the Smartphone by Dr Charakter App

Data Category (No of Types)	Examples of Data	Definition and Description
Standard Android Information (99)	device security, screen layout, colour mode, font scale, keyboard parameters, battery level, rotation set up, Internet set up, alarm alert, bluetooth, tone type when dialing, mute streams affected, vibrate when ringing, etc	current phone settings (during the test)
Call logs (4)	timestamp , number ID, type, duration	Calls data collected from device call history. All phone numbers are anonymised (replaced by random ID)
Text message logs (15)	timestamp , id, type (for sms), average for: number of words, word length, emoticon, MMS, etc	text message data collected from device messages inbox. Data processed on user device. Most common „handmade" emoticons counted.
Contacts (33)	Numbers with COE, with address, with e-mail, contacted last month, with name included family names from the list, etc	List of contacts described in statistics. Number of contacts grouped according 33 different criteria
Pictures (18)	number of photos, average HSV, RGB, dominant colours, width and height, faces, sharpness, rename of photo, etc	Descriptive statistics of users photos on the phone, numbers in directories and characteristics of randomly chosen photos
Exifs (23)	id, obfuscated path, model, long date-time, dimension, orientation, exposure time, flash, focal length, GPS altitude and latitude, re-named by user, photo date taken	Data for photo exifs were collected for all photos without selection
Applications list (5)	package name, names, categories, found url	application data - general information such as the application list
Applications statistics (6)	Period, package name, first timestamp last timestamp, last time used, total time foreground	application statistics - general information such as the date of instalment
Application usage (4)	package name, event type, timestamp, class name	For each installed application – events log this is a big repository of historical information about information usage

(Source: own work)

- **Contact Access** to list of contacts
- **Call Logs Access** to the history of Call logs (last 18 months)
- **Text Message Access** to the list of text message logs (last 18 months)
- **Multimedia** to Multimedia directory containing all camera and photo content (photos and exifs)
- **Application** to Application info, statistics and logs of events (history of usage).

This data typology, dependent on the type of consent granted, is essential for model building. In addition, mobile applications are subject to specific rules aimed at data protection and reduction of data processing only to the necessary data. The model must be based on the data to which this service will have access following its intended use. Therefore, models calculating a personality profile from data should be limited to the data accessed by a given service by definition, and they cannot use the data that is unnecessary from the point of view of the service.

In addition, the issue of the speed of creating the statistics themselves should be mentioned. The data collected and sent to the server was in the form of anonymised data. The data were devoid of identifying marks related to personal data (names, surnames, telephone id, IP, names of photos or catalogues given by the user, content of messages). All the statistics from the phone data, except for photos and exifs, were counted when completing the personality survey and additional questions. Some of the most interesting of them were displayed at the end of the interview to the user. *Dr Charakter*'s research application calculated 18 parameters for the photo and 33 for the exifs. The production tests showed that, depending on the phone model and processor used, the number of photos over 1500 became problematic. For example, for 5000 photos, the calculations could take more than 6 hours or fail because the phone's operating system (Android) closed the DR Charakter application to maintain the system's functioning according to the quality criteria assumed by the manufacturer. The most time consuming was the CV algorithms, which searched for and counted the presence of faces in the photos and calculated the quality parameters of the photos.

The review of the literature showed that the parameters of the photos might be necessary for determining personality (e.g. the number of photos with faces for extraversion and the quality parameters of photos for Openness to Experience), a random mechanism of reducing the photo material for analysis was used in case of more than

1500 photos on the device. The draw procedure used in Dr Charakter App is described in detail in Appendix D. The procedure consisted of stratified layered samples from individual catalogues arranged according to their creation date. The number of samples drawn took into account the structure and number of photos in individual catalogues. The reduction procedure concerned only the calculation of statistics for photos. Statistics from exifs were calculated for the entire tested material (without the reduction mechanism). One of the basic assumptions for using the model was calculating the profile from the moment of installation. The speed of access to the data on which the model is based is a pivotal factor. Considering this, the analysis of the photo material seems problematic, but on the other hand, the speed with which the processing power of processors and the parameters of graphics cards is increased makes one optimistic that this is a problem that will be solved soon.

Descriptive statistics of the collected data and information on qualifying data for analyses will be discussed in the following sections.

3.5.2.1 Data Structure and Characteristic

The collected data is described in detail in Annex 4 in the report on the implementation of data collection. However, some issues require further presentation and comment. The application was installed and research conditions (consents) were accepted by by 2,684 and 2,666 people decided to participate in the study, including 1,303 men and 1,364 women. The histogram for the age groups is shown in the graphic 3.3.

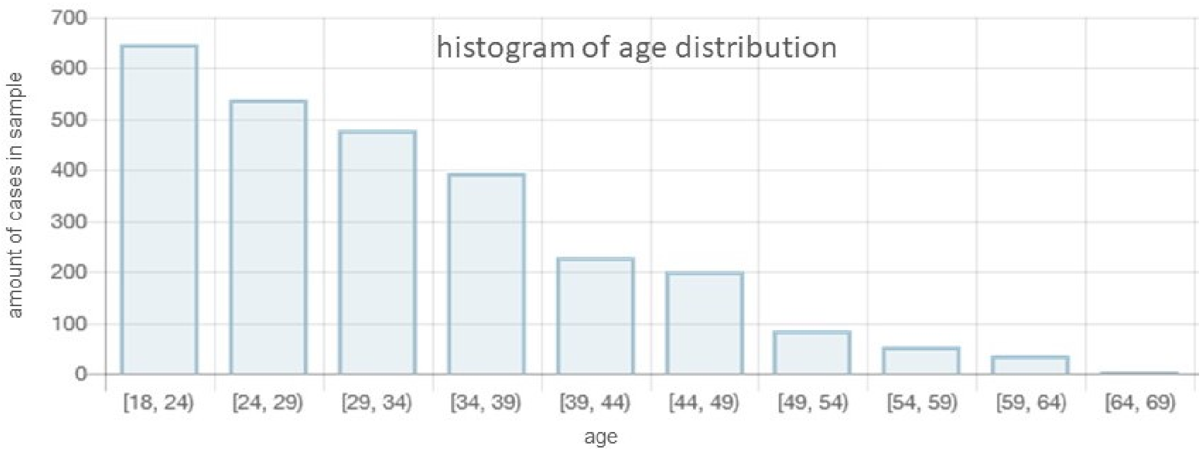


Figure 3.3. Age Distribution - Histogram (own source)

Although the age characteristics are similar to the characteristics of Internet users in Poland, this sample should not be considered representative. It seems reasonable to define it as a target group of people interested in new technologies and possessing technical skills sufficient to independently install the application on the phone and carry out the research by performing commands on the screen. It is not without significance that five consents to access the data required by the application have been granted. These two factors are the critical factor for such age distribution.

On the next histograms, other parameters of the key characteristics of the data used for modelling are presented. All these graphs show how much the data are differentiated between people. There are: a histogram for number of installed apps - Figure 3.4, a histogram for number of calls in call logs - Figure 3.5, a histogram for number of text messages - Figure 3.6 and a histogram for number of exifs - Figure 3.7. An the amount of exifs reports show better the exact quantity of photos stored in the smartphone.

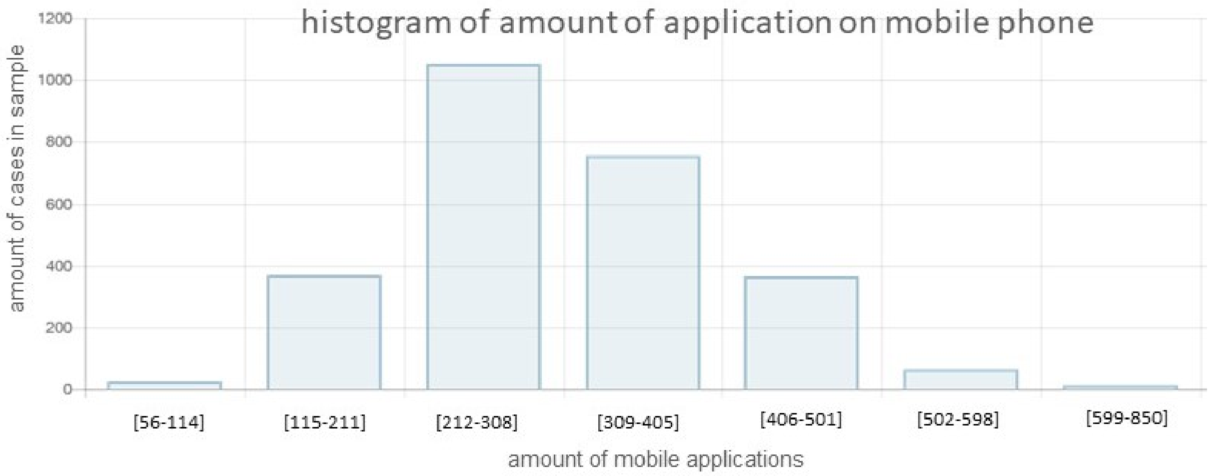


Figure 3.4. Number of Installed Apps - Histogram (own source)

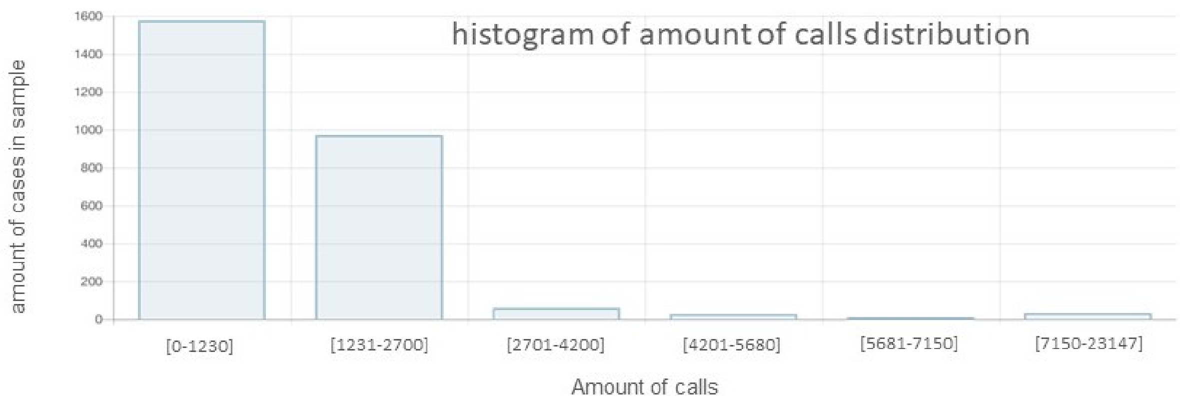


Figure 3.5. Number of Calls - Histogram (own source)

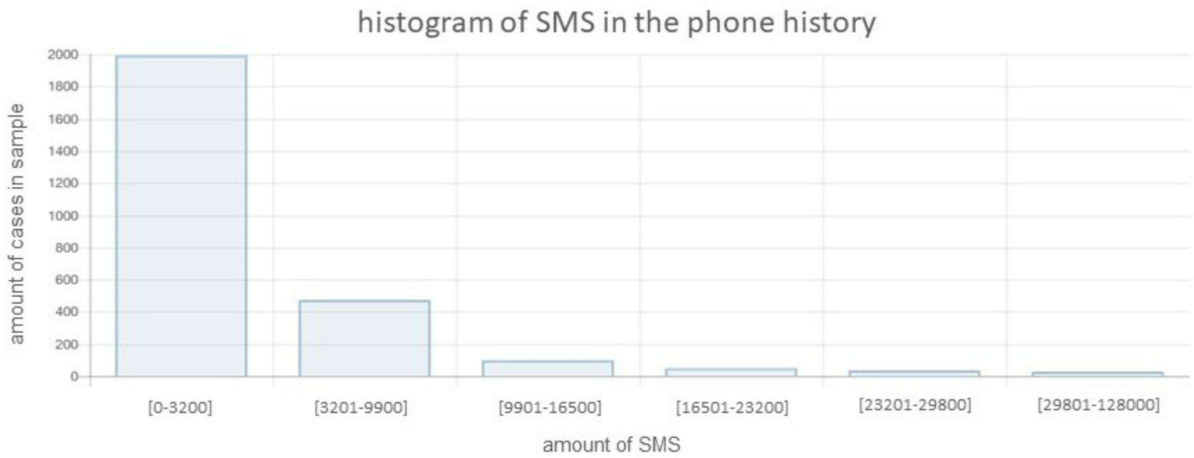


Figure 3.6. Number of Text Messages - Histogram (own source)

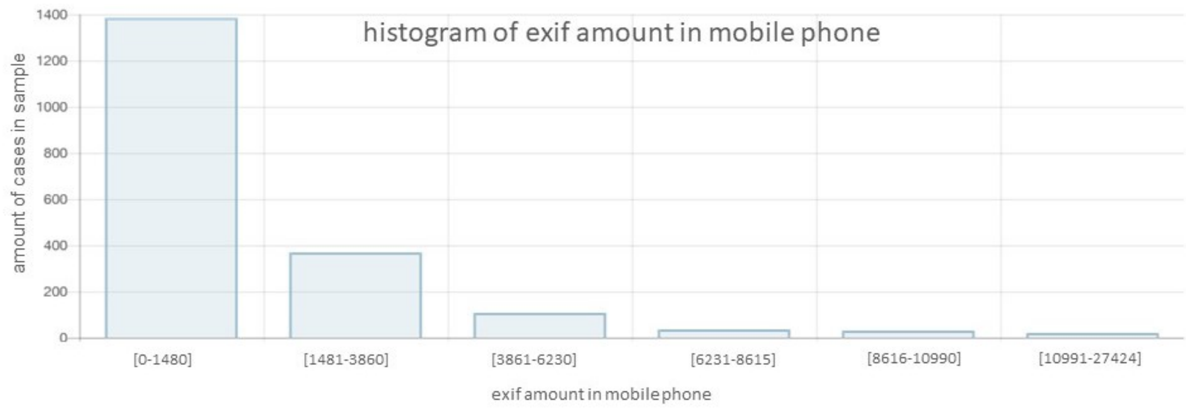


Figure 3.7. Number of Exifs - Histogram (own source)

Additionally basic descriptive statistics are presented in the table 3.3

Table 3.3. Descriptive statistics of the Data Categories Taken from the Smartphone by Dr Charakter App

category of data	number of surveys	average	range (min.-max.)
Number of Contact	2203	174	0-3038
Number of Call logs	2203	1229	0-23147
Number of SMS logs	2203	3178	0-128222
Number of Exifs	1947	1364	0-27427
Number of Apps	2203	51	2-392

(Source: own work)

As can be seen from the Table 3.3, some participants (paid part of data collection) installed applications on clean phones - which probably means they had clean phones for testing (0 as a minimum in the calls, contacts, text messages, exifs categories). These research participants will be excluded at the next stage of the data cleaning process). Additionally, the average amount of exifs is 1485, which means that the picture analysis in the Dr Charakter research app, up to 1500, gives us the analyses from the total number of photos for more than half of the studied sample. The value of 2 as a minimum for the application installed means that this is the lowest number of apps installed on the smartphone by the user. These numbers do not consist of the application installed by a producer (approximately 50 applications pre-installed by handset producer). In the following subsection, the criteria for evaluating the data quality will be discussed.

3.5.2.2 Criteria of Validating Data

In the case of data collected by the *Dr Charakter application*, two types of quality testing or data control were applied: investigation of correctness and coherency of the data declared by a user and data completeness errors. The first study involved looking for people who misplaced the personality questionnaire or installed applications on unused phones. The second involved checking the degree of data that were incomplete or doubled due to the events of interrupted procedures (data transmission errors). Table 3.4 shows the detailed results of these two types of data quality testing and present the final number of verified cases.

Table 3.4. Reasons for excluding cases and the % of initial sample reduction

reason of excluding	number of excluded cases	total number after cleaning	% of initial sample (2666)
data incomplete	58	2608	97.82
time stamp filter	94	2514	94.42
cheating (questionnaire)	17	2497	93.66
cheating (empty phone)	20	2477	92.91
doubled phone data	274	2203	82.63

(Source: own work)

As a result of these two studies, some cases suspected of fraud or insufficiently complete were selected, and ultimately the set for model building contained data from 2,203 people or 82.63% of the collected data.

3.5.2.3 Normalisation of Psychometric Data

Normalisation as a statistical procedure is a standard method of establishing the relationship between the results obtained using a psychometric tool (e.g. test, questionnaire) and the mean results in a given population or, as in the case of these studies, in a research sample (Hornowska, 2018; Magnusson, 1991). As mentioned initially, the tool used in the study to measure the personality profile is not intended to be a psychological diagnosis tool in the sense of a medical diagnosis. The purpose of the standard normalisation to the Sten scale was to determine what range of results can be considered low, average and high. As can be seen from the Table 3.5 containing descriptive statistics for raw data calculated from the questionnaire responses - the distributions differ significantly. The normalisation of the results to the Sten scale will allow the maintenance of the exact definition criteria for all scales and eliminate the influence of the measurement tool itself (perception and reception of the formulations of the test items).

	E	C	A	S	O
Mean	49.6	61.7	72.3	43.7	58.1
SD	21.	18.4	14.5	20.7	15.3
min (0%)	0.0	0.0	3.3	0.0	0.0
25%	36.7	50.0	63.3	30.0	46.7
50%	50.0	63.3	73.3	43.3	56.7
75%	63.3	76.7	83.3	56.7	66.7
max	100	100	100	100	100

Table 3.5. Description of Personality Dimension Distribution, N=3333

(Source: own work)

Normalisation is achieved by calculating how far each raw score obtained by the psychometric tool deviates from the group mean value. Normalisation is, therefore, a procedure that relativises an individual's result to the distribution of the results in the whole group. For model building, normalisation was carried out on a combined sample of all obtained questionnaire measurements, i.e. 3,333 people. The researcher then calculates how far each raw score obtained with a given tool deviates from the mean value in a given normalisation subgroup. The predetermined properties of the normalised scale make it possible to determine what place in the normalisation sample is occupied by a specific result obtained by a given person. For the selected Sten scale, the unit of the standardised Sten scale is one sten. The number of units is 10 (hence the name is an abbreviation of standard ten). It covers 0.5 standard deviations of the population (reference groups), and the mean of this scale is 5.5. Other properties of the sten scale are presented in Table 3.6.

The following formula expresses the transformation that allows converting the raw

Distribution after transformation	normal ('stepped')
Mean	5.5 (between 5th and 6th sten)
Scale jump	1 sten
Standard Deviation	2
Number of scale units	10 stens (range 1-10)
Differentiation	values from -2.25 to +2.25 normalised units 'z'

Table 3.6. Main Properties of Sten Scale according (Hornowska, 2018)

data to the Sten scale:

S = 5,5 + 2 * Z

Where Z is the result of the so-called standardisation of Z (Hornowska, 2018). The advantage of the sten scale is that, firstly, it indicates an approximate person's position relative to other people in that population (each value in the set represents the results of a single person).

Secondly, this individual score is relativised to the normal distribution, allowing for easy determination of 68% of the population with an average score (+ - 1SD from the population mean). Finally, the table 3.7 illustrates the dependency between the values' distribution for normal distribution (Z-scores), Percentages, Percentiles and Stens. This was the base for defining the High, Medium and Low class.

Table 3.7. Standard Z scores, percentages, percentiles, and sten scores comparison (Hornowska, 2018)

Z-scores	<2	2to1.5	1.5to1	1to0.5	0.5to0	0to0.5	0.5to1	1to1.5	1.5to2	>2
Percent	2.28%	4.41%	9.18%	14.99%	19.15%	19.15%	14.99%	9.18%	4.41%	2.28%
Percentile	1.4	4.48	11.27	23.36	40.43	59.57	76.64	88.73	95.52	98.86
Sten	1	2	3	4	5	6	7	8	9	10

The histogram set (Figure 3.8) compares the distributions for the 5 personality dimensions. Before normalisation, these are raw results converted from questionnaire responses plotted on a 100-point scale (the same as the% of possible points). After normalisation to the Sten scale, it is the size distribution for each of the ten created classes. As in the case of other questionnaire measurements of personality (Strus and Cieciuch, 2014), the raw data already shows that the distributions of Extraversion, Openness to Experience and Emotional Stability are similar to the normal distribution.

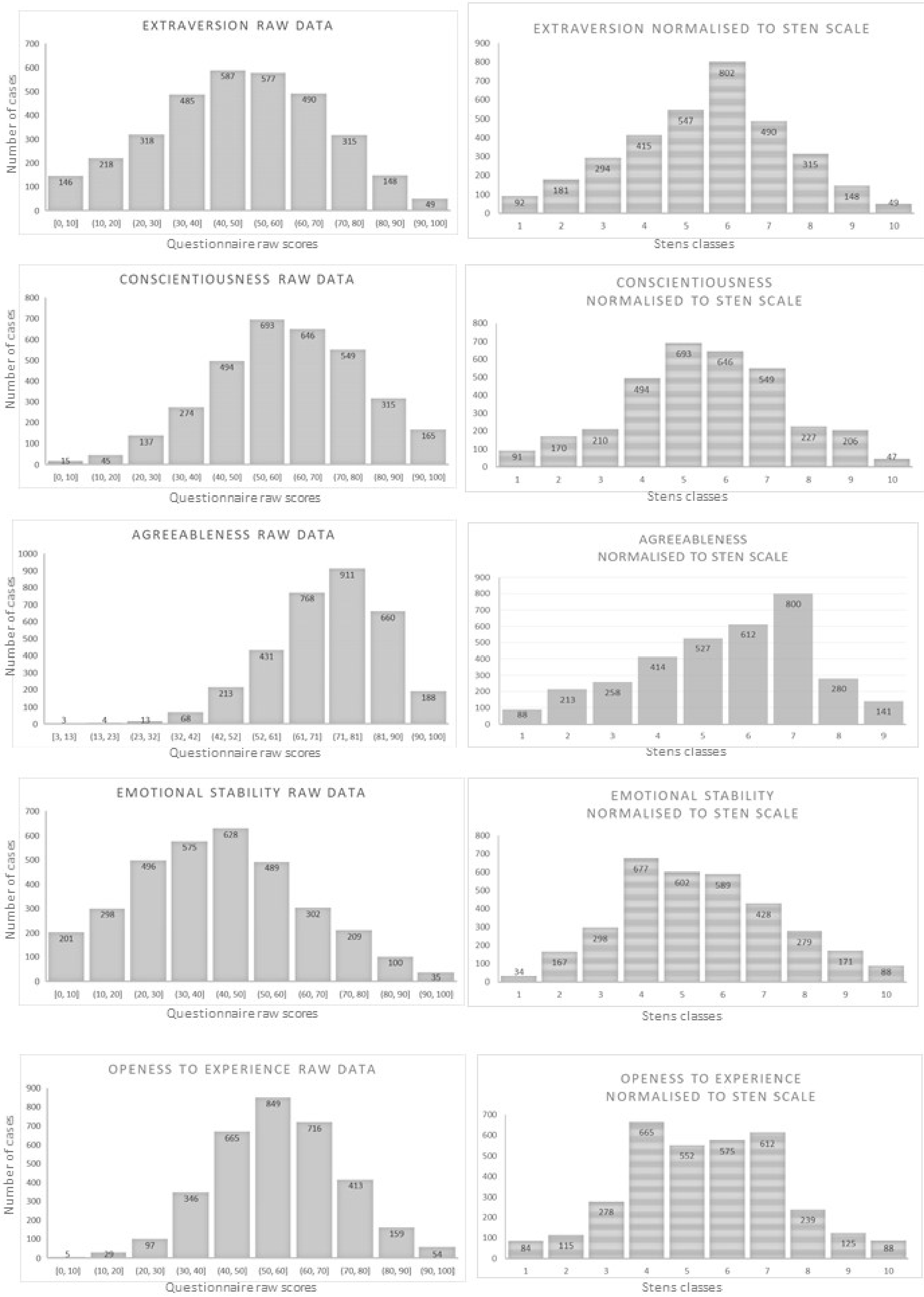


Figure 3.8. Comparison of Personality Dimensions Distribution before and after the Normalisation - Histograms

(Source: own work)

The Conscientiousness and Agreeableness distributions are skewed (histograms in Figure 3.8). For raw data distribution, skewness and kurtosis range from -0.87 to 0.99 and have the highest values for Agreeableness. It is probably the subject of a separate research program to investigate what influences the personality dimension obtained from the questionnaire tool. In the case of having more empirical material, it would be scientifically valuable to repeat the above analyses to study phenomena repeatable among psychometric research. The research hypothesis about the cause of the skewness for the Agreeableness dimension requires further investigation. It can result from tool conditions (lack of resistance to social bias: the need for social approval, low self-esteem).

However, the reason can also lie in need for presenting oneelf more beneficially) or be an effect of the volunteer method of data gathering for example. From a psychological perspective, there is a higher probability to have High Agreeable people in the sample of volunteers, additionally in task connected with personal data sharing. So all data gathered on volunteers are biased in the same way. This fact can also be treated as a potential limitations for this and comparable research. Such bias must also be considered in the methods of personalisation based on the data-driven personality.

3.5.3 Summary of the Data Used for Modelling Review

Chapter 3 summarised all preparatory stages connected with the defined artefact creation. The research requires creating dedicated tools like Big 5 psychometric inventory and an Android application for processing and gathering statistics from a mobile phone. During the extensive literature review (chapter 2), it can not be found any comparative researches.

Together with the conclusion from the literature review, it seems that a unique set of diversified data used for modelling is the most vital point of research. Firstly, such a data set allows for comparing the predictive power of various digital data in connection with personality. Secondly, the data gathering procedure makes it possible to conduct personal analytics and modelling on a fully anonymised data set.

The use of different recruitment methods for the study (volunteers and paid recruitment) and a detailed analysis of the questionnaire filling related behaviours provided additional information on the effectiveness and quality of these two recruiting methods. The discussed limitations of these methods should always be carefully considered due to their impact on test results. In addition, it is worth having natural results of the personality questionnaires together with statistics, which allows their combining and aggregating, re-classifying or possible iterative feature engineering.

Despite data being valuable for this and further research, there is also space for improvement. Additional analyses connected with the scale and quality of feature engineering during the modelling process can bring crucial changes to how the data are pre-processing (creating anonymised statistics) on the phone. For example, a review of existing research has concluded that photographic data can be critical in personality prediction. However, the processing of photos on the end device caused the most problems when collecting data. In addition, creating new variables that would have a chance to differentiate the behaviour of individuals requires different methods and the processing of other variables.

Additionally, preliminary examination of the photographic data proved unsatisfactory. Contrary to the data from the application and the system, which significantly impact personality prediction, this type of statistics should be expanded and explored further in subsequent studies.

The last but not least point for discussion is the data timeliness in the context of the standardisation function of the model. Undoubtedly, additional research is required to develop methods of data relevance assessing in the context of rapidly changing factors such as consumer behaviour, mobile phone use profile, telephone software, and data availability. Based on comparative data from different points in time, it will be possible to determine how often the model should be updated (recalculated). Behaviours related to modern devices such as smartphones are closely related to the functions offered by the devices. An example is a massive change of calling and sending text messages using the telco network into calls and sending messages using solutions based on data transmission such as Messanger, WhatsUp or Hangout (Google Chat).

Chapter 4

Proposed Method of Using Data-Driven Personality Model

The previous chapter described the research scheme and procedure of collecting data and data pre-processing, necessary for calculating the personality profile based on smartphone data (stage 6 and 7 in the diagram of research steps in the Figure 1.3). The current chapter is devoted to artefacts creation (step 8 of the research). It consists of two main parts. The first is appropriate for creating a user's initial smartphone personality profile, the UISPP model (Artefact 1). The second section describes personality based automatic personalization of app, named the PAPA method (Artefact 2). Finally, a discussion about the limitations of the research and solution is performed.

4.1 General Concept Defining

In the literature review, Section 2.3.2. presents qualitative criteria for the evaluation of studies for building personality models based on the data proposed by (Azucar et al., 2018). Azucar et al. defined three universal criteria: data were related to personality scores on an individual (personal) level, personality should be measured based on a standardized and validated personality measurement tool (not necessarily diagnostic), and user data had to be collected in an automated manner. Chapter 3 presents a detailed description of the entire research procedure, the purpose of which was to collect data appropriate for the construction of the model. It was also shown that all three criteria were met in this dissertation: the psychometric procedure and psychometric parameters for the personality measurement tool were presented, and data was collected in an automated manner using the Dr Charakter application, thanks to which anonymous data from the personality survey was combined with smartphone statistics. One of the purposes of this dissertation is to create a personalisation method with complete protection of the privacy of data and user profiles. Therefore, these studies must use only fully anonymous data and statistics, with full transparency for their collection, automatically resolving discussions about the ethical use of personal data and sensitive data for scientific purposes. As a rule, data used to calculate the profile as well as the profile itself, should be treated as sensitive data and must be strictly protected. This implies that these records may not be used without the consent of person concerned (Finck, 2021).

As discussed earlier (mainly in Chapter 2.6), some critical business requirements are associated with the proposed solution. They are directly related to the gaps identified in the literature review concerning the existing solutions. These currently do not provide the possibility of automated personalisation based on the created user profile from the moment of installing an electronic service such as a smartphone application. In addition, they were confirmed in qualitative research with users, who also pointed out the disadvantages of existing solutions and problems with profiling based on the history of events in the account. These requirements are, in a sense, also limitations of the proposed solution. Summarising:

1. The personality profile must be calculated on the user's end device (smartphone) to make the full data protection possible and, as a result, the availability of the profile should be limited to the profile's owner. This condition directly relates to the gap identified in the literature review (Ad point 2, 4, 5 and 6 in Section 2.6)
2. The profile must be calculated from the available data at the time of service installation so that automatic profiling of the service can take place from the start. In the case of a mobile application, this must be done shortly after installing the application on the phone and receiving the consent of the user. This condition refers to the gap identified in the literature review (Ad points 1, 3, 4 in Section 2.6)
3. There is no need to have a detailed personality profile for initial profiling. It is enough to estimate each trait on a low-medium-high and, in this way, define per-

sonality dominants (traits on high and low levels). For personalisation purposes, those who have special uncommon needs and preferences are need to be identify. This condition directly relates to the gap identified in the literature review (Ad points 7 and 8 in Section 2.6)

4. As various personalised elements of the service are adjusted independently to the level of individual personality dimensions, potentially, the entire profile does not have to be counted for all services. Models will be created independently for each dimension and can also be used independently, for example only for service execution differentiation for Introverts and Extraverts. In this case, there is no need to count the entire profile (all five dimensions). Such approach allows for elimination of redundant non-essential user information and increase of technical efficiency parameters. This condition refers to the gap identified in the literature review (Ad points 4, 8 in Section 2.6).

Keeping in mind the defined gap based on literature review, as well as the business requirements and conditions, the following artefacts were defined and developed:

- **(A.1)** A method of determining a User's Initial Personality Profile based on the minimum amount of data available at the time of installation (UISPP).
- **(A.2)** A method of implementing Personality based Automatic Personalisation of App (PAPA) based on UISPP model.

The fulfillment of these conditions is necessary for artefact 1 and artefact 2 to make a significant contribution to the existing knowledge base. The entire process of creating and evaluating artifacts is subordinated to these requirements. In addition, the building and development of items will take place in laboratory conditions, in the development phase of test mobile application for UX/UI surveys.

Although the UISPP model (A.1) is based on the well-known and widely used ML analytics, it is performed on new data categories (smartphone data) that have not yet been explored. The solution's innovation also improves the method of calculating the user's personality profile and enabling it to be counted in a short time on the end device without the need to transfer data to external IT systems. The implementation of this goal enables complete protection of the user's data and profile; no data is disclosed outside the customer's end device (smartphone) and remains confidential. By design, it meets all the requirements of the GDPR and eliminates the problem of the lack of protection of customer information that is created based on data. This execution is the consequence of requirements defined for PAPA method (A.2) implementation dedicated for any mobile app personalisation from the moment of installation. The creation of the PAPA method is interdependent with the creation of the UISPP (A.1). The dependence is presented in the Figure 4.1, collecting determines the order in which artifacts are developed.

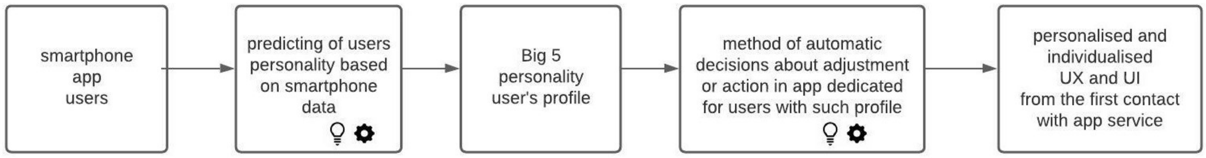


Figure 4.1. General Concept of Proposed Solution

(Source: own work)

To sum up, the proposed solution aims to identify people with unusual needs and adapt an electronic service such as a smartphone application to them. The key aspect of this solution is the full automation of such service personalisation and necessity of collecting behavioural data from the service (related to identified gap points 1, 3 and 4 in Section 2.6). In addition, despite the use of sensitive data about the user, the data and information created on their basis can remain inaccessible to the service provider. Therefore information about user cannot be used for purposes other than auto-personalisation of the service offered to him/her. The innovative nature of the solution is influenced by the fact that the entire process is automated and sewn into the customer's end device.

4.2 Development of Data-Driven Personality Model

The main goal is to create an algorithm that calculates the personality profile to identify groups of users with different needs, using the data available on the user's phone at the time of installing the application. Accessing some of the smartphone data requires obtaining various consents (including general approval of data processing). This is why depending on the permissions granted by the user, different data sets will be obtained for model creation (data and related consents are described in Section 3.5.2). The User's

Personality Aware Service idea of automatic self-personalising to individual's needs (for example, related to the need for stimulation (Extraversion)) from the service installation is presented in Chapter 2.6.

According to the chosen research methodology (see section 1.3), after the problem identification (Step 1) described in Chapter 1, a literature review related to the problem (Step 2) described in Chapter 2 was carried out. Then, the initial research with potential Users (Step 3) described in Appendix A and summarized in subsection 3.3 was was completed with by the design and development of research tools. These exclusively created tools made it possible to conduct dedicated studies to collect the required data (Steps 4, 5 and 6). Chapter 4 describes in detail the process of data analysis and artefact creation.

In Chapter 1 of this dissertation, the artefacts are defined. Then, in Section 4.3, the creation of the first artefact will be presented (Artefact1: A method of determining a user's initial personality profile based on the minimum amount of data available at the time of service installation). Then, the creation of Artefact 2: A method of implementing a personality profile for automatic personalisation of the service (based on example of mobile app) is presented in Chapter 4.4.

4.2.1 User's Initial Smartphone Personality Profile (UISPP) Model

This section of the dissertation is dedicated to developing the artefact 1: User's Initial Smartphone Personality Profile Model (UISPP). It is a predictive model, the purpose of which is to classify the phone user, based on the data available on the phone, into one of 3 classes on five personality dimensions. In this way, a model of personality will be assumed, and personality dominants as well as preferences for the user's needs regarding the application will be determined. Finally, based on these pre-defined user preference classifier, the application will use the PAPA method to prepare a profiled version of the service (application).

4.2.2 Technical Environment

The UISPP model was created on the standard PC unit (processor: AMD Ryzen 5 PRO 3500U w/ Radeon Vega Mobile Gfx 2.10 GHzwith, with RAM 16,0 GB and 64-bits operational system, processor x64 and Windows 10) using standard open-source Jupyter software in Anaconda environment using Python programming. Most of the computation, including all machine learning techniques and methods, was conducted using the scikit-learn library. This library is commonly used for machine learning for both scientific and business purposes (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, Blondel, Prettenhofer, Weiss, Dubourg, Vanderplas, Passos, Cournapeau, Brucher, Perrot, and Duchesnay, 2011; Raschka, 2018). Scikit-learn library covers many implementations and co-operates reasonably with other Python analytical libraries like NumPy, SciPy, and matplotlib. Training of the models, search grids and other automatic tasks takes no longer than a few hours. Comparable models can be created without any special technical equipment.

The dataset gathered by the Dr Charakter app are presented in Section 3.5, and the complete list of variables linked with user consent needed are included in Appendix C.

4.2.2.1 Machine Learning Techniques Applied

Machine learning (ML) is one of the branches of analytics related to artificial intelligence (AI), which involves the learning of machines (computers) without the need for direct programming (Samuel, 1959). Thanks to used learning algorithms, computers can independently analyze data and automatically adapt models to changing phenomena and requirements to acquire new knowledge and increase their ability to solve a given problem. For example, machine learning grading uses a mathematically proven algorithm guide to perform complex and computationally requiring analytical tasks. Moreover, the suitable algorithms and an adequately trained model assure the accuracy classification that humans could never achieve. The workflow applied in this research includes the following main steps:

- Step1: Exploratory Data Analysis and Feature Engineering;
- Step2: training and testing different kinds of classification algorithms separately in the context of the model for five personality dimensions;
- Step3: exploring different methods of algorithms aggregation;
- Step4: utilising some components of Auto ML;
- Step5: comparison of results and creating final UISPP with fine-tuning hyperparameters.

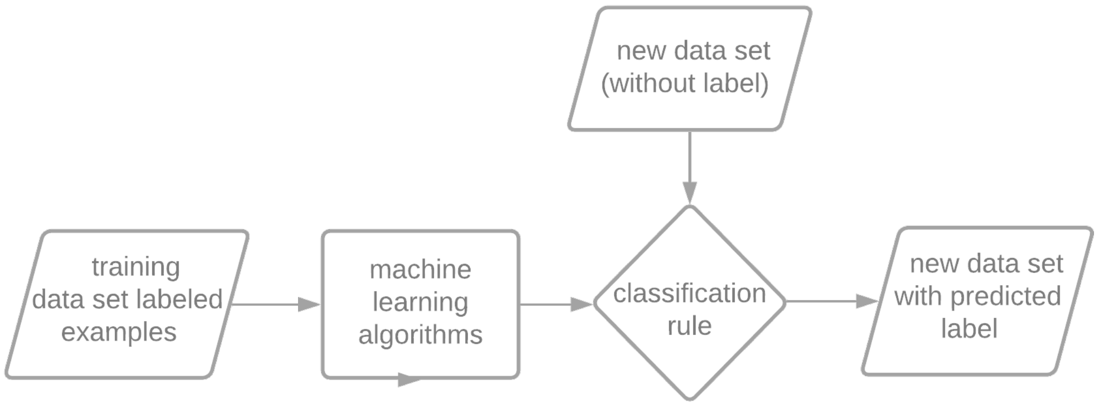


Figure 4.2. Machine Learning General Workflow

(Source: own work)

Simplistically, ML leads to creating a model that will be a complex rule of classification for newly acquired data. This simplified diagram is shown in the 4.2. Based on iterative learning done on the training data set, an algorithm is created that will enable the creation of a classification rule independently of the initial data (horizontal course in 4.2). In turn, this kind of universal rule, applied to a new data set with the same characteristics, will make it possible to predict the class sought with some specific validity (vertical course in 4.2). To build a model that will determine the user profile based on the available data (when installing the service on a smartphone), the verification of several available ready-made machine learning algorithms was used. The sections below describe the steps of building this model in turn and selecting the final method based on the quality parameters of the model. This methodology is beneficial when fast classification/prediction is required for real-time decision making, such as service personalisation. Statistical (or mathematical) methods are used to derive a model from the observed data set without assuming specific guidelines or model defining commands (Bishop, 2006).

There are 12 ML predicting methods chosen among others for tests. All of the algorithms used, listed in the table 4.1, are supervised ones. The supervised learning algorithms try to model relationships and dependencies between the personality dimension and the input features. Based on those relationships, the output values for new data can be predicted with some probability. What is essential, as is clear from the Big 5 theory, the dimensions are independent, which means that for getting a complete

Table 4.1. Comparison of different machine learning algorithms

Algorithm	Description
Random Forest Classifier (RF)	an ensemble method usually used for classification and regression problems, method is based on the construction of a multitude of decision trees during training,the outcome is a dominant class or can be a mean prediction from many trees.
k Nearest Neighbour (KNN)	a supervised technique of classification which uses proximity as a proxy for similarity. It takes labelled points and uses them to learn how to label other points.
Support Vector Machine (SVM)	supervised learning models with associated learning algorithms used for classification and regression issues.
Logistic Regression (LR)	finding the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable) and a set of independent (predictor) variables.
Linear Discriminant Analysis (LDA)	is a generalisation of Fisher's linear discriminant, a method used to find a linear combination of features that characterises classes of objects or events.
Gaussian Naive Bayes (NB)	classification technique based on Bayes' Theorem with the assumption of independence among predictors
Extra Trees Classifier (ETC)	implements a meta estimator that fits many randomised decision trees on various dataset sub-samples of and uses averaging to control over-fitting and improve the predictive accuracy.
Bagging Classifier (BC)	an ensemble meta-estimator that fits base classifiers on random subsets of the original dataset and then aggregates their predictions (either by voting or by averaging) to form a final prediction.
Decision Tree Classifier (CART)	the technique which builds classification or regression models in the form of a tree structure; works by dividing a data set into smaller subsets while, at the same time, a decision tree is incrementally developed.
AdaBoost Classifier (ADA)	type of meta-estimator which at the beginning learns to fit a classifier into the original dataset and then fits additional copies of the classifier into the same dataset but where incorrectly classified weights of instances are adjusted in such a way that subsequent classifiers focus more on complex cases.
LGBM Classifier (LGBM)	a gradient boosting framework that uses tree-based learning algorithms. LightGBM grows tree vertically while other tree-based learning algorithms grow trees horizontally.
XGB Classifier (XGB)	an optimised distributed gradient boosting library; the algorithm implements machine learning techniques under the Gradient Boosting framework.

Big 5 profile, it is required to build five independent models, one for each dimension. In each of them, it will be predicted if a person belongs to the high, middle or low class in each of the five traits (E, C, A, S, O). At the preliminary analysis stage, an approach was also tested in which ten models are created, i.e. a separate model for predicting a high score of the personality dimension and a separate model for predicting a low value of a given dimension. However, the predictions based on this approach turned out to be less accurate, and thus further work was abandoned.

4.2.2.2 Ensemble Learning: Gradient Boosting Machines and Model Combination

Combining multiple algorithms or machine learning models, known as ensemble techniques, combines base models to create one heterogeneous predictive model with better quality parameters. This procedure provides better model stability (reduces the randomness of the prediction) and increases the efficiency of the prediction. The ensemble methods aim to find the heterogeneous model that best predicts the user's personality class. The ensemble methods are an alternative to the complex selection of one best model. By design, no model is reliable in all cases. Therefore, team methods consider the results of many models and average them to obtain one final model. The most common method is decision trees. Collaborative techniques such as the highly parallel bootstrap aggregation meta-algorithm (also known as bagging) are used in algorithms such as random forests to average the predictions of individual decision trees while effectively reducing over-fit. Improving qualitative parameters is iterative and gradually adjusts to weak learners, e.g. decision trees with pre-constraints. In turn, increasing the gradient is done in adaptive type algorithms such as ADA, adding gradient descent elements, building new models, and optimising the cost function different from the errors in previous iterations. Based on the findings from the literature review (Section 2.3.2.2) XGB and LGBM are the two most used gradient ML algorithms after 2016, differing in their tree-building logic. While creating artefact 1, both algorithms based on ensemble methods such as Bagging and Boosting (RF, BC, ADA, XGB, LGBM), as well as advanced algorithm stacking techniques, such as majority voting (Hard Voting) and weighted average (Soft Voting) were used.

4.2.2.3 The Problem of Skewed Class Label Distribution

In the case of creating models predicting the personality profile, we are dealing with a prediction of extreme levels of traits with dispersal close to the normal distribution. This

Table 4.2. The number and distribution of groups defined as low and medium high categories for each of the 5 traits in the collected data.

	Low (1-3 sten)	Medium (4-7 sten)	High (8-10 sten)	Low (1-3 sten)	Medium (4-7 sten)	High (8-10 sten)	Total
EXTRAVERSION	358	1595	249	16%	72%	11%	2202
AGREEABELNESS	402	1547	253	18%	70%	11%	2202
CONSCIENTIOUSNESS	286	1630	286	13%	74%	13%	2202
STABILITY	296	1581	325	13%	72%	15%	2202
OPENNESS	314	1683	205	14%	76%	9%	2202

is related to the problem of imbalanced class label distributions. Such label distribution distortions or class imbalances can lead to ineffective predictions as models can optimise the learning outcome by learning to predict the most abundant label. The table 4.2 presents the numbers and shares of predicted groups.

Standard methods such as *'train-test-split'* do single-layer sampling by default, which can result in class label distributions not being adequately represented compared to the original data set. Consequently, there is a risk of over-fitting. To prevent this, the *'train-test-split'* tiered sampling option has been used so that the class labels in each resulting sample match the distribution existing in the given input data set. This method eliminates the problem of the small number of classes in the test and training samples, but it does not prevent preferring the label of the most represented class in the population in the forecast. As a result of the imbalanced classification, there are too few cases of minority classes for the model to successfully 'understand' the decision boundary, which prevents efficient prediction. In the case of personality, such a majority class represents ambivalent results, i.e. average results, constituting about 68% of the population according to the normal distribution. The problem was solved by using several techniques to balance the distribution of class labels. imbalanced learning can be solved with four techniques to support the gradual differentiation of classes.

There are two techniques of data re-sampling, either reducing the number of occurrences of data samples that contribute to class over-representation (under-sampling) or generating new data samples from underrepresented classes (over-sampling) (Huda et al., 2018; Tyagi and Mittal, 2020). Generating new data samples is about creating artificial data, not copying existing cases from the minority class. Therefore, improving the material for building the prediction consists in enlarging the collection of examples from the minority class by artificially creating new examples from the minority class. Since over-sampling tends to train models that exceed the data, the third technique combines oversampling with the technique of 'cleaning up' under-sampling, removing extreme outliers in the majority class. The fourth technique is related to the ADA model, in which the model set is built from different insufficiently tested major class sets, and the minority classes are trained separately. This technique allows the use of more data from the over-represented class as an alternative to mere re-sampling.

The applied Synthetic Minority Over-sampling Technique (SMOTE) is a general approach to the synthesis of new examples. SMOTE has been described by Chawla et al. (2002). The technique is to select examples that are close to the function space by drawing a virtual line between the examples in the function space and selecting a new point lying on the line between the existing cases. As the first step, a random example from the minority class is selected. Next, KNN is chosen, and then a random neighbour is selected, and a synthetic example is created at a randomly selected point between the two examples in the feature space. In this way, new cases are generated precisely in the spaces existing in the set of actual cases.

4.2.3 UISPP Model Construction

In creating a UISPP model, the input data is the data available on the user's phone, and the output data is the result class for one of the Big 5 personality dimensions. The general concept of this process is illustrated in 4.3. Learning outcomes may vary depending on the specifics of the problem, the nature of the data, and the methods. Therefore, creation of a model is carried out based on the training-testing-verification cycle of multiple repetitions. In the following steps, the use of other method and variable configurations is verified, the techniques are confirmed by the performance quality indicators as discussed in chapter 2.3.2, Recall, Precision, F1 and Accuracy. Iterative training is sometimes embedded in the method itself, such as in auto ML, and sometimes requires manual adjustment of parameters. As mentioned, the basis for personalisation will be the predictive model (UISPP), which will be used in the service (smartphone application) using the PAPA method. Machine learning aims to find the best model to correctly predict the affiliation to one of the three classes (Low, Medium, High) for each of the five personality dimensions (E, C, A, S, O). This means the UISPP will consist of 5 different models. In practice, the number of models may go up to over 50 if the number of data consents is considered a condition for data availability. The automatic processes of

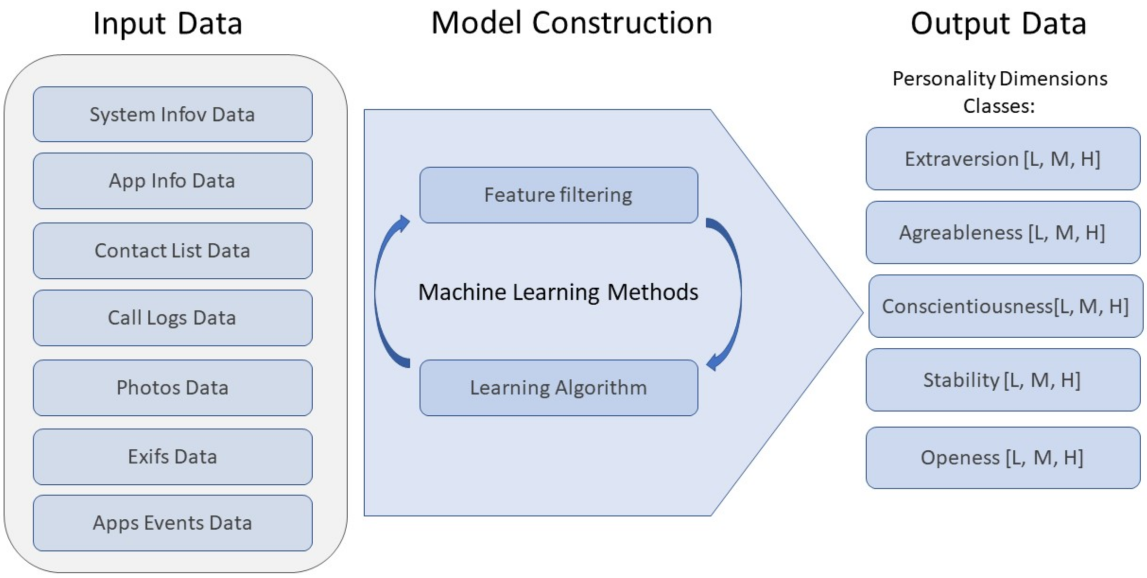


Figure 4.3. The General Concept of Machine Learning to Create a UISPP Model

(Source: own work)

the qualitative assessment of models are aimed at maximising the Accuracy parameter. Since it is practically impossible to implement a 100% accurate model for predicting complex behaviours, reference levels should be defined for model evaluation. In the next chapter, 4.3.4, the strategy of recognising the model as valuable by exceeding baselines determined from various reference points will be discussed.

4.2.4 Defining Baseline for Model

During the machine learning process, the model creation process evaluates the iteratively created models regarding their usefulness for forecasting. The most common method of assessing model usability is to define a baseline forecast. If the models created in the machine learning process are better than this baseline forecast, this is the basis for considering that the model is valid and whether its creation brings added value. It also allows for quick evaluation and rejection of some tested prognostic techniques and focuses on improving those algorithms that predict the best. There are several methods for calculating a base score.

```
Random for  -1 0 1 - conscientiousness_123
              precision    recall  f1-score   support

          -1       0.15      0.37      0.21       286
           0       0.76      0.35      0.48      1630
           1       0.13      0.34      0.19       286

    accuracy                           0.35      2202
   macro avg       0.35      0.35      0.29      2202
weighted avg       0.60      0.35      0.41      2202

Random for -1 0 1 - extraversion_123
              precision    recall  f1-score   support

          -1       0.16      0.32      0.22       358
           0       0.72      0.34      0.47      1595
           1       0.12      0.35      0.18       249

    accuracy                           0.34      2202
   macro avg       0.34      0.34      0.29      2202
weighted avg       0.56      0.34      0.39      2202

Random for -1 0 1 - agreeableness_123
              precision    recall  f1-score   support

          -1       0.18      0.30      0.22       402
           0       0.70      0.34      0.46      1547
           1       0.11      0.34      0.17       253

    accuracy                           0.34      2202
   macro avg       0.33      0.33      0.28      2202
weighted avg       0.54      0.34      0.38      2202

Random for -1 0 1 - stability_123
              precision    recall  f1-score   support

          -1       0.15      0.34      0.20       296
           0       0.74      0.35      0.48      1581
           1       0.15      0.34      0.21       325

    accuracy                           0.35      2202
   macro avg       0.34      0.35      0.30      2202
weighted avg       0.57      0.35      0.40      2202

Random for -1 0 1 - intellect_123
              precision    recall  f1-score   support

          -1       0.14      0.30      0.19       314
           0       0.78      0.35      0.49      1683
           1       0.10      0.35      0.15       205

    accuracy                           0.35      2202
   macro avg       0.34      0.34      0.28      2202
weighted avg       0.63      0.35      0.41      2202
```

Figure 4.4. Screen for Baseline Models defined for Random Prediction Model (Randomisation Coherent with Each Trait Distribution)

(Source: own work)

```
Medium level  (0) for everyone - conscientiousness_123
              precision    recall  f1-score   support

          -1       0.00      0.00      0.00       286
           0       0.74      1.00      0.85      1630
           1       0.00      0.00      0.00       286

    accuracy                           0.74      2202
   macro avg       0.25      0.33      0.28      2202
weighted avg       0.55      0.74      0.63      2202

Medium level (0) for everyone - extraversion_123
              precision    recall  f1-score   support

          -1       0.00      0.00      0.00       358
           0       0.72      1.00      0.84      1595
           1       0.00      0.00      0.00       249

    accuracy                           0.72      2202
   macro avg       0.24      0.33      0.28      2202
weighted avg       0.52      0.72      0.61      2202

Medium level (0) for everyone - agreeableness_123
              precision    recall  f1-score   support

          -1       0.00      0.00      0.00       402
           0       0.70      1.00      0.83      1547
           1       0.00      0.00      0.00       253

    accuracy                           0.70      2202
   macro avg       0.23      0.33      0.28      2202
weighted avg       0.49      0.70      0.58      2202

Medium level (0) for everyone - stability_123
              precision    recall  f1-score   support

          -1       0.00      0.00      0.00       296
           0       0.72      1.00      0.84      1581
           1       0.00      0.00      0.00       325

    accuracy                           0.72      2202
   macro avg       0.24      0.33      0.28      2202
weighted avg       0.52      0.72      0.60      2202

Medium level (0) for everyone - intellect_123
              precision    recall  f1-score   support

          -1       0.00      0.00      0.00       314
           0       0.76      1.00      0.87      1683
           1       0.00      0.00      0.00       205

    accuracy                           0.76      2202
   macro avg       0.25      0.33      0.29      2202
weighted avg       0.58      0.76      0.66      2202
```

Figure 4.5. Screen for Baseline Models calculated for Non Personalised Service (The Business Perspective)

(Source: own work)

According to the previous considerations (Chapter 2), the constructed predictive model should classify service users for each of the five personality model traits into levels of trait values with the values ’Low’ / ’Medium’ / ’High’, determined based on the result of the normalised sten scale (see Section 3.5.2.3):

- Low level - sten 1 to 3
- Average level - sten from 4 to 7
- High level - stens from 8 to 10

Based on model usage, three different approaches to creating the base parameters for the model were pre-analysed and considered.

Random Perspective: The first approach is a baseline created building on 3 class randomisation for the scenario, taking into account their distribution in the population established on the distribution of labels in the data set collected to create the model. Considering that, in the case of a classification problem, the base level will be the most common class - the modal of levels (named Average or Medium level; stens from 4 to 7). For individual personality traits, founded on the distribution of classes (Low, Medium, High), calculated baseline levels of performance metrics are as presented in Figure: 4.4.

Personalisation Perspective: The second is the approach resulting from the business goal, i.e. using the model to personalise the service. In this approach, the benchmarks are the quality parameters obtained in the absence of personalisation, i.e. a scenario in which only the modal class is ’correctly identified’ because only it receives a typical product tailored to his/her non-specific needs. Calculated baseline levels of performance metrics are presented in the the Figure: 4.5

Benchmark Research Perspective: The third approach is to calculate the qualitative parameter based on existing research on building similar models presented in Section 2.3.2. (Table 4.3). While this is the only benchmark available from existing studies that aimed to build a personality model from data, it has its limitations. First according to the best knowledge, no studies calculate personality models based on static data available on a smartphone at a given moment, and most are based on text data from social media. Thus it is a completely different research material. Second, in most published studies, there was no prediction of 3 classes but two binary ones. There are more differences between individual studies, but the two essential ones apply to everyone. It is not easy to estimate how the differences between the research relationship affect the obtained quality parameters. The range between the results of one study, on the same data, models for different dimensions of personality, is sometimes significant: e.g. 0.45-0.75 (Golbeck et al., 2011), 0.45-0.57 (Kalimeri et al., 2013).

Table 4.3. Mean Accuracy for Benchmark Baseline

Description of the ase for calculating Accuracy	Accuracy
The single highest Accuracy achieved among 14 papers (a case of Hexaco model from eye-tracking)	0.85
The mean from the highest achieved Accuracy from all 14 papers in which Accuracy was reported	0.690
The mean from the average Accuracy, from all 14 papers in which Accuracy was reported	0.643
The mean of the highest Accuracy from 8 papers in which the Big 5 was predicted	0.636
The mean average Accuracy from 8 papers in which the Big 5 was predicted	0.608

(Source: own work)

The Table 4.3 shows the calculated mean accuracy values for the test results grouped according to different criteria. The highest accuracy (0.85) was found in the study, where personality was predicted based on tracking eye-movement. The lowest reference value (0.606) is the average accuracy of eight studies in which the Big 5 was determined based on the digital data. The benchmark from the published studies has the limitation that many publications lack detailed data on the residual models for individual features (they reported only average values). The limitations of the research itself (other types of data, personality dimensions as binary variables) do not appear to be fully comparable. Ultimately, however, the best method of predicting a personality profile from the data is sought, so the mere fact of a difference in the type of data or a division into the number of classes can be considered favourable or worse prediction.

Tables 4.4 contains all counted qualitative measures that will be used for the validation of the created models, i.e. Precision, Accuracy and F1-score. Taking into account the accuracy itself, which can be compared with the published studies, it can be seen that the reference ranges from the studies (0.45 -0.85) significantly exceed the level of accuracy in random baseline (0.34-0.35). That difference means that regardless of the data used for prediction and the method, it is possible to determine the model that predicts the personality profile above the random baseline. The situation is different in the case of a baseline created according to the business approach related to personalisation (4.5). Here, the range for accuracy alone is significantly higher (0.70 for A to 0.76 for O). Only a personality model based on eye-tracking exceeds this baseline - creating a higher level for comparison. The average of the highest levels for studies determining the Big 5 profile (0.71) is within the range determined by the business baseline (0.70-0.76).

Finally, it can be concluded that the baseline determined randomly is inadequate if existing attempts to create similar models are taken into account (state of the art benchmarks). Furthermore, for most features, the business baseline, defined in terms of increasing relevance for personalisation, seems to be more ambitious than the highest relevance for existing research. Thus, considering external benchmarks from existing research will be used in assessment and evaluation.

4.2.5 Preprocessing Feature Engineering for the Model (Step 1)

Data preprocessing and feature engineering is the initial process of constructing the appropriate features from the available data to improve predictive performance. Feature engineering is based on arithmetic, aggregating, and relativising transformations to generate new variables having better discriminative and predictive properties. Initial data preprocessing was described in Chapter 3, involves excluding data incomplete cases and outliers. Feature engineering is an art closely related to domain knowledge rather than a structured process. The whole process was carried out in two stages. The first can be described as the creative phase, and the second is the automated phase. Considering data from the phone, automatic techniques for examining data sets were not adequate in the initial phase due to their diversity. Initially, it was manual work within each variable category. The techniques used are scaling, normalisation, time limitation, relativisation of continuous data, i.e. standardisation. When looking for the distinguishing features of behaviour, standardisation is the most common procedure already at the stage of creating new variables. New features are created by generating relationships between two variables, which become a better discriminant of human preferences and behaviours. For example, in calculating the profile of a single person, the description is created by relativisation. Instead of the total number of installed applications or number of educational applications, a new variable is created: the ratio of educational applications to all applications. In this way the set of 261 features was cre-

Table 4.4. Share of features with Mutual Information higher than 0 for each personality dimension and features' category

	Number of features	Total	E	A	O	S	C
aplication list	113	43%	40%	39%	34%	45%	45%
calls history	43	16%	19%	20%	18%	17%	16%
contact list	30	11%	15%	9%	12%	14%	13%
phone settings	75	29%	26%	31%	36%	23%	27%
	261	100%	100%	100%	100%	100%	100%

ated (the complete list is available in Appendix 5). The Table 4.4 presents the numbers of input variables used for modelling for each data category. The researcher's creativity concerning the defining of variables is the only limitation of generating the new feature.

Mutual Information (MI) is a statistic for measuring the degree of relatedness between data sets. MI detects relationships between data sets, whether they are mean values or variances or extremes. MI is interpreted as the amount of information exchanged between data sets (measured, for example, in bits). The Mutual Information method is based on information entropy theory and is a non-linear method of measuring accuracy and redundancy between pairs of variables. Entropy is used as a measure of uncertainty for each variable, and MI is this information enumeration in two variables (Nargesian et al., 2017). For example, the MI between two data sets, X and Y, can be estimated from the two data sets' pairwise (x, y) statistics. Thus, MI is an effective measure for finding relationships between discrete and continuous variables. Moreover, this is the case with phone data and personality class.

Mutual Information was calculated automatically by commands integrated with the scikit-learn library in Python soft-wear (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, Blondel, Prettenhofer, Weiss, Dubourg, Vanderplas, Passos, Cournapeau, Brucher, Perrot, and Duchesnay, 2011). The number of variables selected based on Mutual Information is presented in table 4.4 and the completed list is presented in Appendix 5. For all attempts of creating models, only these sets of MI features were used.

4.2.6 Comparison of Single Algorithms (Step 2)

After the feature extraction step using the MI method, the first planned step was implemented. Its primary goal was to identify the set of the best-performing algorithms on the collected data set. Twelve algorithms, described in Section 4.3.2.2, has been tested in the classic training-testing model presented in Figure 4.6. In addition, all twelve were checked in the classic Holdout Set by one-time split and in Holdout with Cross Validation. The first step in building a machine learning model is training and testing (Pal and

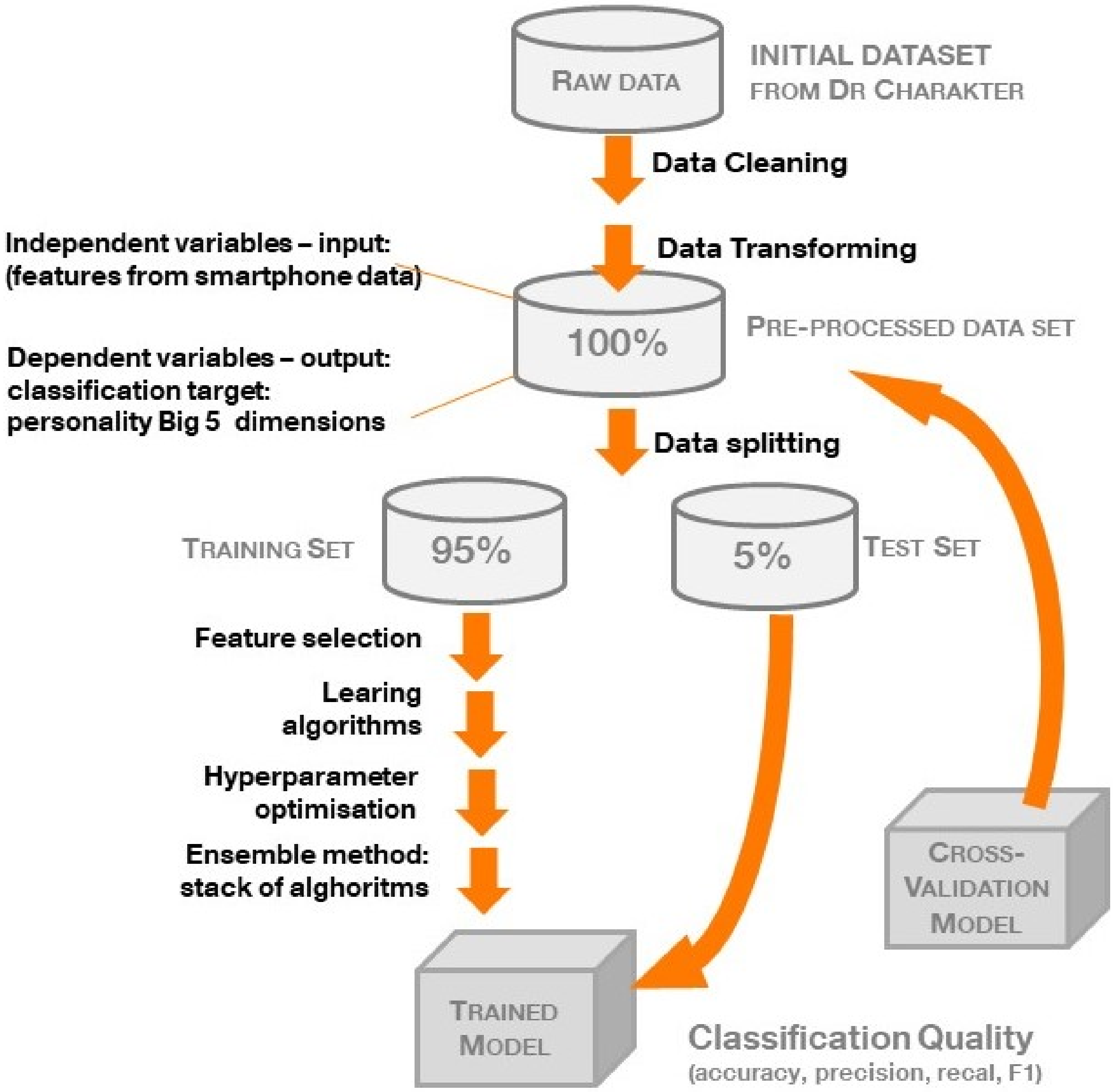


Figure 4.6. The General Steps for Construction the Predictive Model

(Source: own work)

Patel, 2020). Both training and testing start with data partitioning. According to the trait distribution, in the stratified sampling, 5% of the entire set was drawn and set aside for validation, the so-called holdout set. The remaining 95% is a training set, divided into a training set and a test set in a four-to-one ratio (Figure 4.6). The 80% training set is a subsection of the dataset that a machine learning algorithm discovers or from which it 'learns' about the relationship between the smartphone statistics (features) and the personality trait's level. As training data is marked with results in a personality test, learning is supervised.

The remaining 20% is the test subset of the input against which the learned machine learning algorithm is tested to see how accurately it identifies the relationships between the known outcomes for the target variable and other functions of the dataset. The holdout subset, which is also a test set, is used to evaluate the performance of a machine learning model after training and validation. Holdout sets should not include cases used to decide which algorithms to use or improve or fine-tune the algorithms. Providing these conditions enables the creation of appropriate models for the data to be collected in the future, not just the data on which the model was trained.

In the Holdout with Cross-Validation (Figure. 4.7), the same list of algorithms was checked using 5-fold cross-validation. For cross-validation, the default method is 5x cross-validation with 20% latency. The resulting subsets (folds) are used n-fold for training and testing, and the program iteratively turns the folds for training and testing. Thus, there are as many iterations, as there are folds (in the case of the conducted study, five were used). In this way, the result is a robust classifier that performs better on different dataset variations, and its efficiency is not dependent on a random sampling of the data.

Therefore, one experiment is with the different class production techniques used to find the best algorithm for a given data type. The research work objective is to choose the best machine learning technique by experimenting with actual data, dividing it into folds (k number), and observing the results. The generated files are split into nearly equal parts. N times denotes the n iterations of the model on each change of training and test data set. Figure 4.7 shows how splitting data sets works.

In the first step, the features (selected based on MI) are converted into numbers, and a vector space model is created with a numerical representation of the feature and the corresponding number of occurrences. The classification model consists of 2 modules: a training module and a test module. All the techniques used differ in building a model of the relationship between features and personality traits. The testing module checks the learned model with a given method and uses it to predict a personality trait on entirely

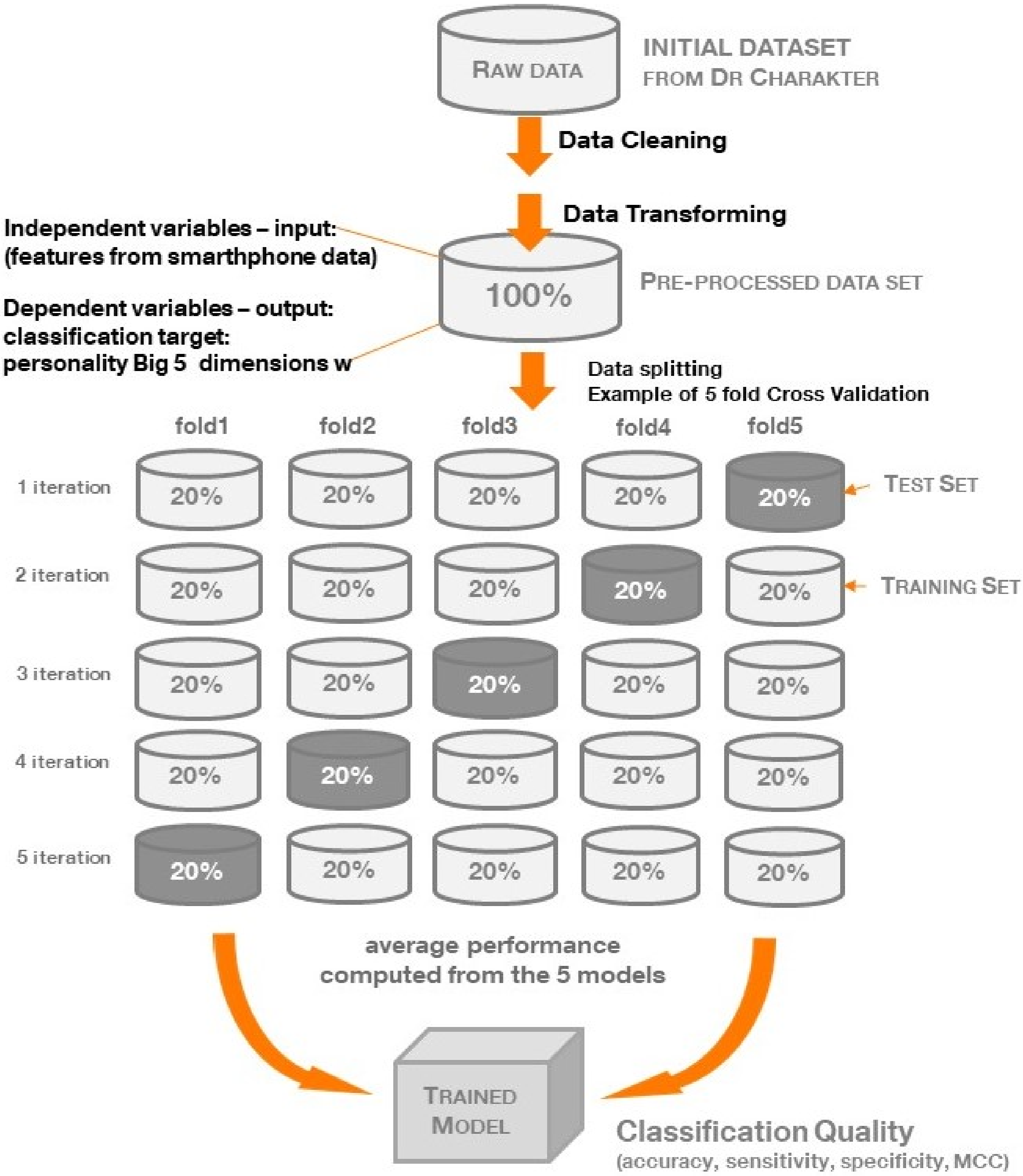


Figure 4.7. The Detailed Steps for Construction the Predictive Model with Cross Validation Scheme

(Source: own work)

new data that was not used to train the model (Figure: 4.6).

The results of this initial analytical step are summarised in tables 4.5 and 4.6. The best-achieved results belong to RF, ETC, LGM, XGB and BC. The range of values obtained for the quality parameters is quite extensive. On the other hand, some algo-

Table 4.5. Comparison of Single Algorithms Performance in classic hold-out set

P for Precision, A for Accuracy, F1 for F1 score - definition explained in Section 2.3.2

	E			**C**			**A**			**S**			**O**		
	P	A	F1	P	A	F1	P	A	F1	P	A	F1	P	A	F1
RF	**.62**	**.67**	**.65**	.55	.72	.62	.58	.67	.58	.55	.69	.60	.58	.71	.64
ETC	**.66**	**.70**	**.68**	**.55**	**.74**	**.63**	.49	.68	.57	**.52**	**.72**	**.60**	**.59**	**.77**	**.67**
BC	.63	.67	.65	.55	.72	.62	**.62**	**.69**	**.60**	.59	.69	.62	**.67**	**.76**	**.68**
LR	.69	.37	.38	.59	.40	.45	.57	.39	.44	.58	.46	.49	.64	.42	.49
LDA	.72	.37	.39	.59	.38	.43	.56	.40	.44	.60	.45	.48	.65	.42	.49
KNN	.59	.39	.44	.60	.28	.30	.58	.28	.31	.64	.28	.30	.54	.20	.19
CART	.60	.52	.55	.59	.53	.56	.54	.49	.51	.55	.48	.51	.70	.66	.68
NB	.38	.11	.40	.79	.20	.11	.38	.13	.09	.61	.22	.19	.35	.17	.10
SVC	.66	.39	.42	.63	.60	.61	.52	.50	.51	.60	.64	.61	.55	.51	.53
ADA	.64	.46	.51	.63	.60	.58	.49	.51	.50	.60	.61	.61	.58	.59	.58
LGBM	**.61**	**.66**	**.63**	.54	.72	.62	.58	.67	.58	**.62**	**.71**	**.63**	**.64**	**.75**	**.68**
XGB	**.63**	**.65**	**.64**	**.55**	**.73**	**.63**	.53	.64	.56	.57	.70	.61	**.68**	**.75**	**.69**

(Source: own work)

rithms obtained unsatisfactory qualitative results below the random baseline, for example, KNC, NB, KNN.

Table 4.6. Comparison of Single Algorithms Accuracy in the Hold Out (HO) dataset with CVCross Validation (CV) dataset

	E		C		A		S		O	
	HO	CV	HO	CV	HO	CV	HO	CV	HO	CV
RF	**.67**	.60	**.72**	**.72**	**.67**	**.67**	**.69**	**.69**	**.71**	**.75**
ETC	**.70**	**.63**	**.74**	**.73**	**.68**	**.69**	**.72**	**.71**	**.77**	**.76**
BC:	**.67**	.55	**.72**	**.63**	**.69**	.58	**.69**	**.62**	**.76**	**.68**
ADA:	.37	.47	.40	.61	.39	.54	.46	.57	.42	**.63**
CART:	.37	.51	.38	.53	.40	.52	.45	.52	.42	.53
LR:	.39	.32	.28	.41	.28	.39	.28	.41	.20	.42
LDA:	.52	.35	.53	.41	.49	.39	.48	.40	**.66**	.43
KNN:	.11	.38	.20	.26	.13	.26	.22	.23	.17	.43
NB:	.39	.18	.60	.18	.50	.16	.64	.19	.51	.17
KNC:	.46	.26	.60	.26	.51	.35	.61	.34	.59	.34
SVC:	**.66**	.35	**.72**	.59	**.67**	.54	**.71**	.58	**.75**	**.64**
LGBM:	**.65**	**.65**	**.73**	**.73**	**.64**	**.69**	**.70**	**.70**	**.75**	**.75**

(Source: own work)

In cross-validation, quality parameters are usually worse. Based on the accuracy of checked algorithms, those chosen for the next step of the model creation were RF, ETC, LGM, XGB and BC.

4.2.7 Model Build on Voting Classifier (Step 3)

In the first step, the performance of individual algorithms on the collected data intended for model learning was checked. Five algorithms that achieved the best results in terms of the adopted qualitative measures (Precision, Accuracy, F1 score) were selected. The next step will be to check whether it is possible to use the dissimilarity of algorithms and then decide about belonging to the predicted personality classes based on the compliance of prediction from different algorithms. A voting classifier is a machine learning model that learns from several models and predicts an outcome (class) based on the highest probability for a selected class as an outcome. The algorithm aggregates the results of each classifier passed to the voting classifier and predicts an output class based on the majority of votes ((H. Wang et al., 2013)). In this way, a single model is created, which learns from reading the predictions of multiple models. There are two voting methods: the hard one and the soft one. Hard is based on choosing the class with the greatest number of votes - the class with the highest probability that should be selected by each classifier (algorithms) used. Suppose three of the algorithms predict the output class for the extraversion trait as (H, H, M). In this situation, the High Extraversion result will be the most likely forecast. Thus, the hard vote is a simple majority vote. In soft voting, the output class is a prediction based on the average probability for that class. Again for 5 algorithms predicting the extraversion dimension assume the example probabilities for the high level H = (0.45, 0.65, 0.35, 0.53, 0.4), for the medium level M = (0.5, 0.3, 0.35, 0.38, 0.45) and for the low level L = (0.2, 0.25, 0.3, 0.2, 0.2). So the average for class H is 0.475 and M is 0.396, and for L = 0.23, the winner will be class H because it had the highest probability averaged by each classifier. While using each of the algorithms separately, the results of classes M, H, M, H, M, respectively, would be obtained. The key seems to be using different algorithms to eliminate any error made by one of the algorithms. Voting classifier helps to make a decision in doubtful situations where the result of the classification varies depending on the algorithm that is used.

4.2.7.1 Hard Voting Classifier Outcome

The graphs below present the averaged qualitative results for individual algorithms and the result obtained due to the Hard Voting classifier. There are five charts: one for each personality trait. Extraversion: Figure 4.8, Conscientiousness: Figure 4.9, Agreeableness: Figure 4.10, Stability: Figure 4.11 Openness to Experience: Figure 4.12.

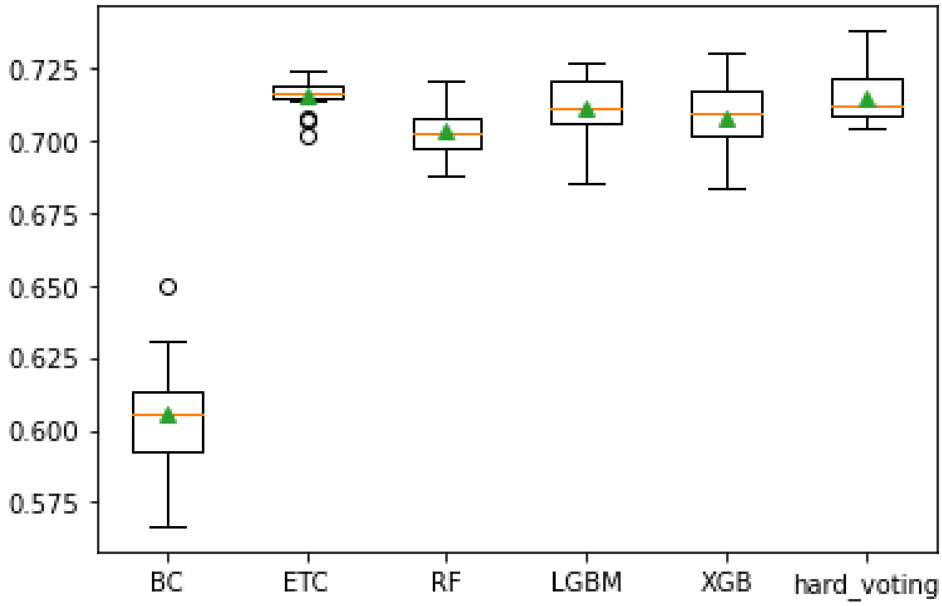


Figure 4.8. Hard Voting Results Comparison of Accuracy for Extraversion

(Source: own work)

Stacking algorithms with a Hard Voting classifier is a procedure of majority voting. In this way, the final prediction is equal to the class achieving the highest number of votes. Based on comparing the accuracy of all five dimensions prediction, the BC algorithm performs the worst. Others (ETC, RF, LGBM, XGB) achieved comparable results, and stacking does not significantly improve accuracy.

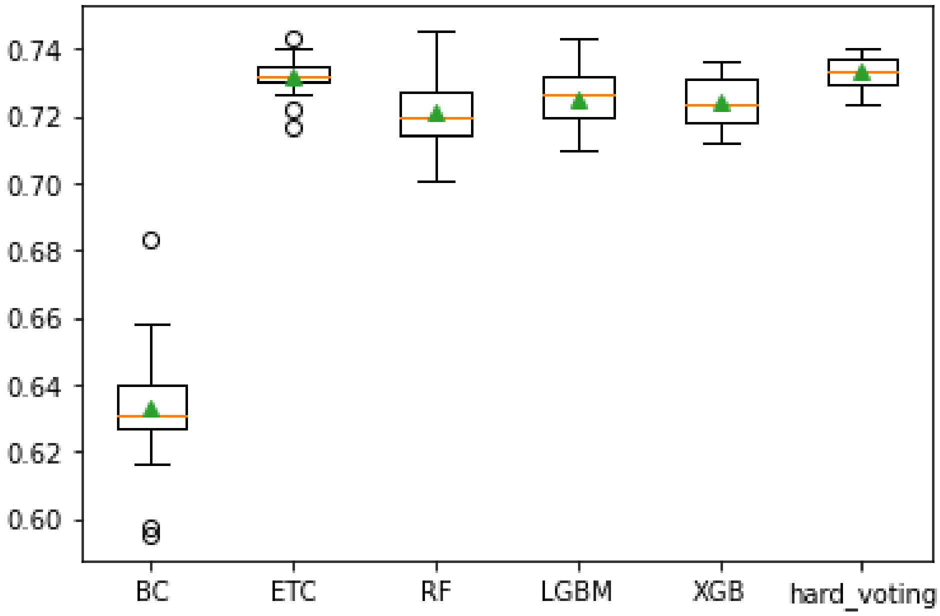


Figure 4.9. Hard Voting Results Comparison of Accuracy for Conscientiousness

(Source: own work)

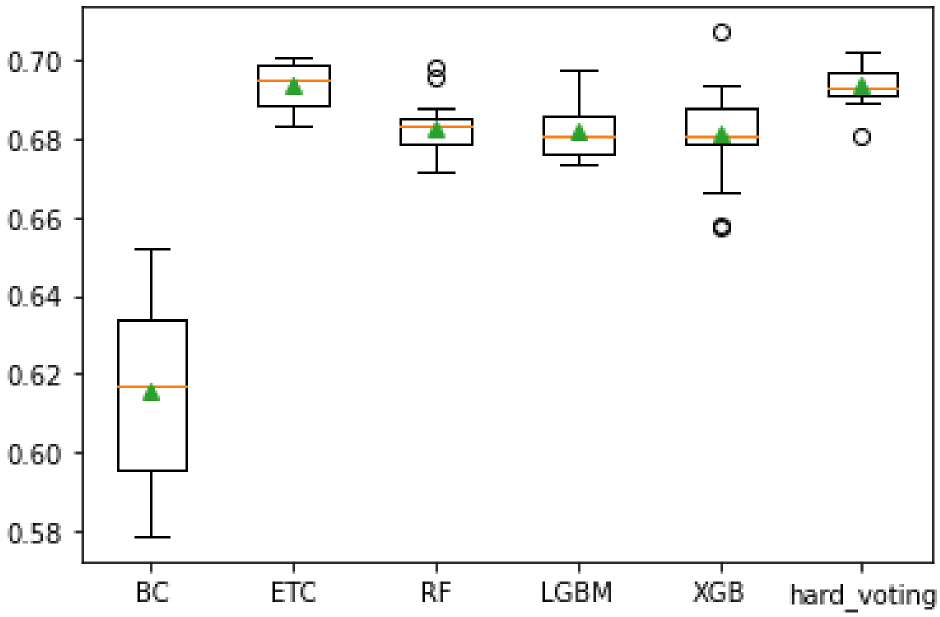


Figure 4.10. Hard Voting Results Comparison of Accuracy for Agreeableness

(Source: own work)

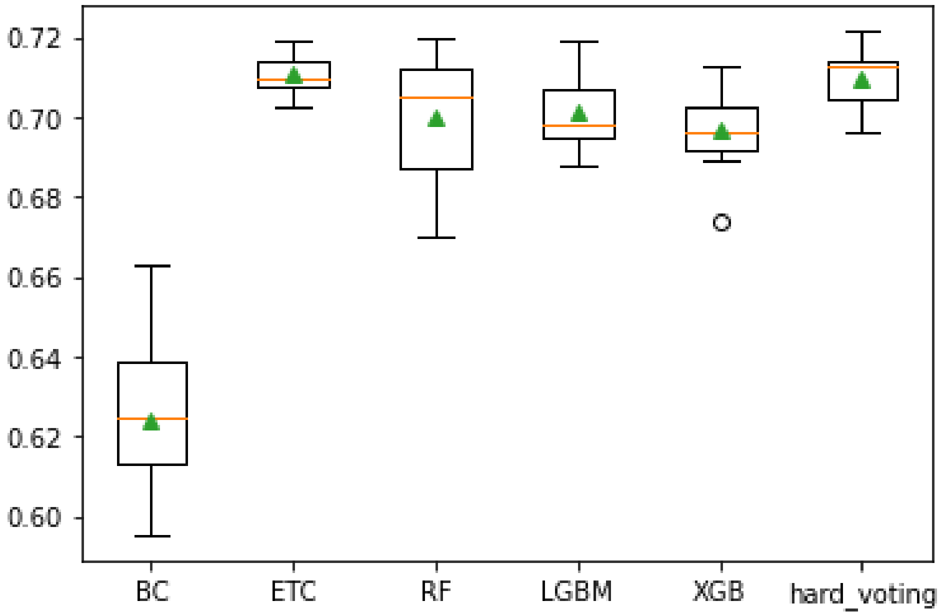


Figure 4.11. Hard Voting Results Comparison of Accuracy for Stability

(Source: own work)

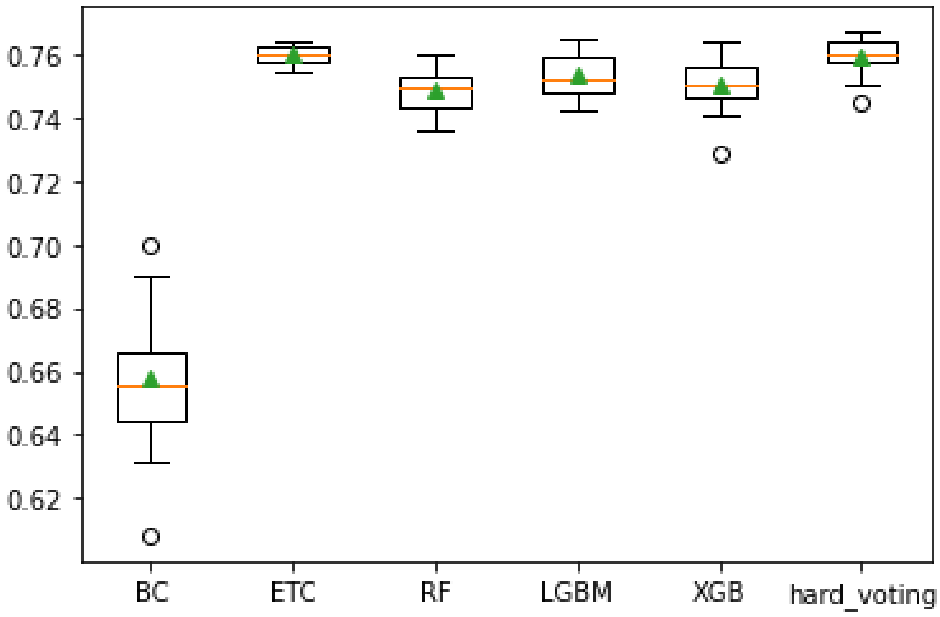


Figure 4.12. Hard Voting Results Comparison of Accuracy for Openness

(Source: own work)

4.2.7.2 Soft Voting Classifier Outcome

The graphs below present the averaged qualitative results for individual algorithms and the result obtained due to the Soft Voting classifier. There are five charts: one for each personality trait. Extraversion: Figure 4.13, Conscientiousness: Figure 4.14, Agreeableness: Figure 4.15, Stability: Figure 4.16 Openness to Experience: Figure 4.17.

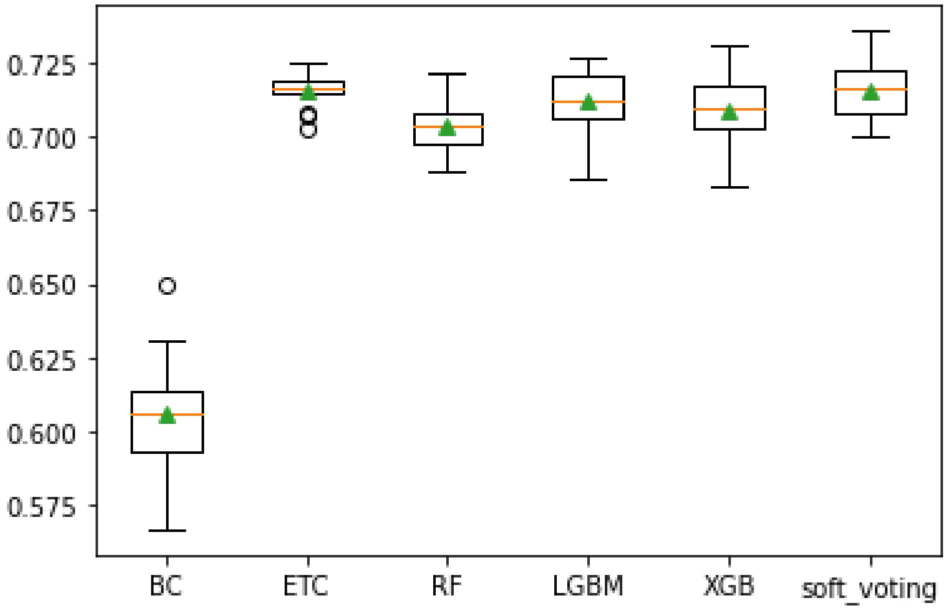


Figure 4.13. Soft Voting Results Comparison of Accuracy for Extraversion

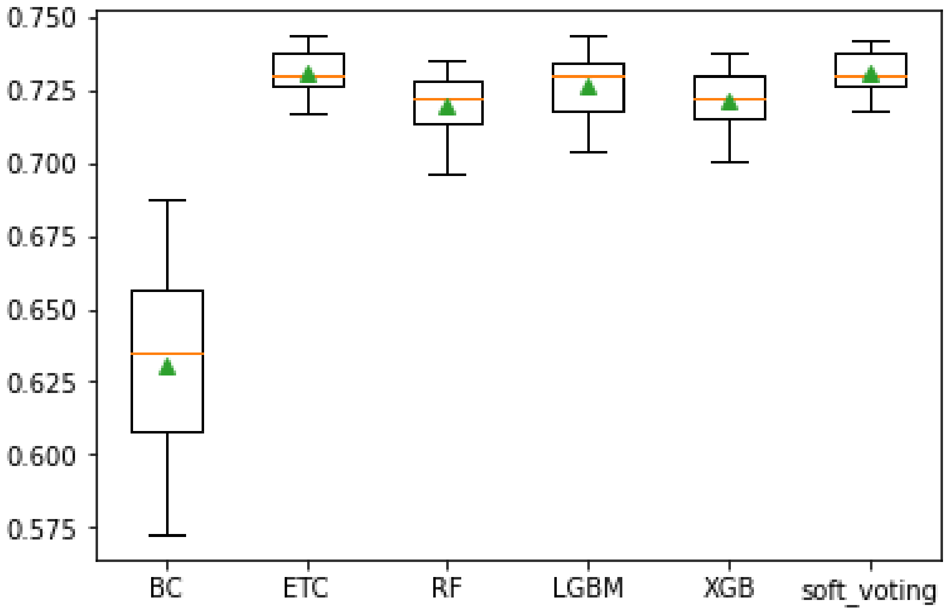


Figure 4.14. Soft Voting Results Comparison of Accuracy for Conscientiousness

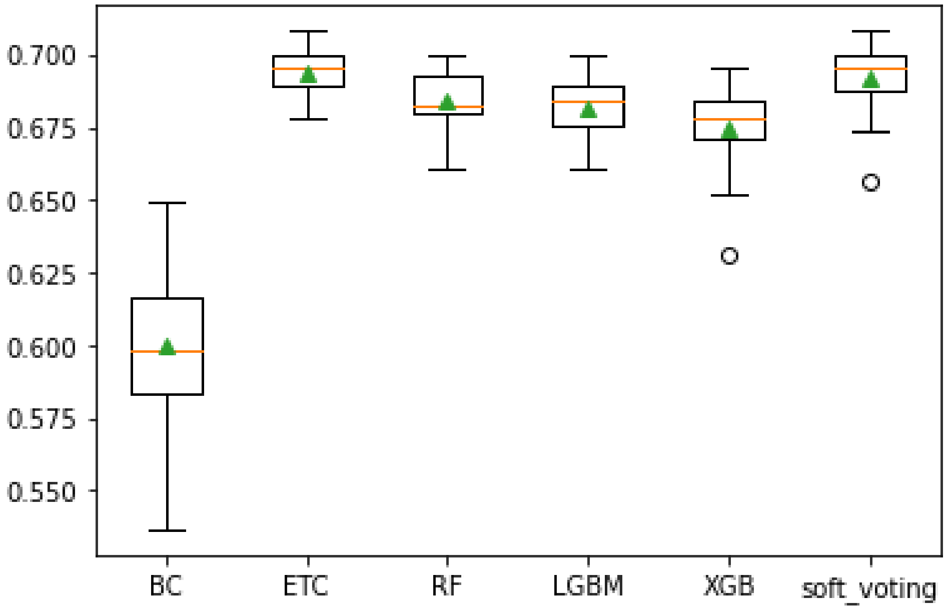


Figure 4.15. Soft Voting Results Comparison of Accuracy for Agreeableness

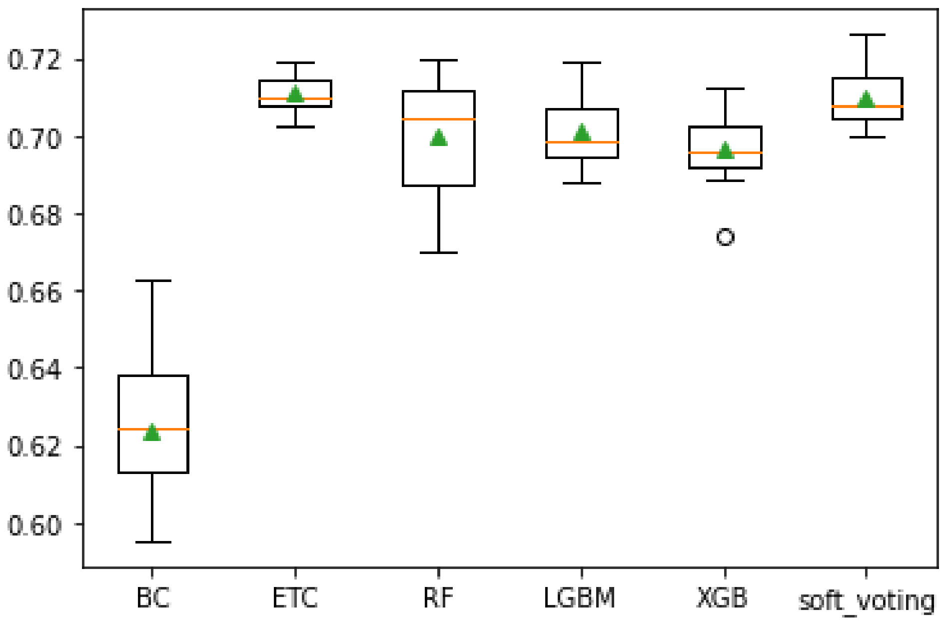


Figure 4.16. Soft Voting Results Comparison of Accuracy for Stability

Stacking algorithms with a Soft Voting classifier is a procedure of weighted voting. In the first step of soft voting, each classifier gives a value for the probability that a specific person belongs to one of the target classes. In the second step, the forecasts are weighted according to algorithm importance and summed up. The final label is the one with the highest sum of weighted probabilities.

Based on comparing the accuracy of all five dimensions prediction, similarly to in the Hard Voting experiment, the BC algorithm performs the worst. Others (ETC, RF,

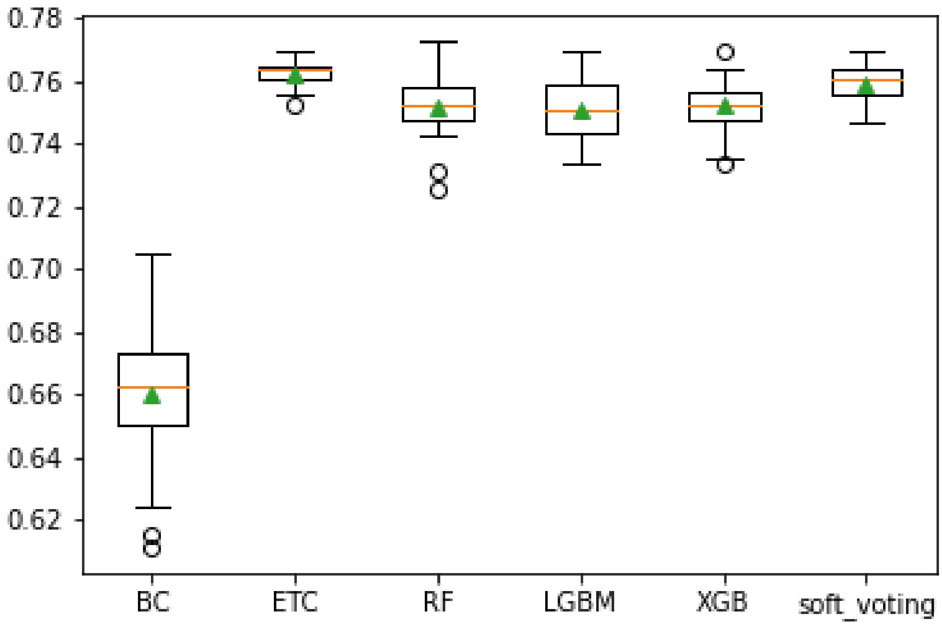


Figure 4.17. Soft Voting Results Comparison of Accuracy for Openness

LGBM, XGB) achieved comparable results, and soft voting stacking does not significantly improve accuracy.

To sum up, the experiments with the use of the best stacking algorithms through Hard and Soft Voting did not bring any improvement in the quality of prediction. The percentage of correctly predicted classes is in the same range of values for the stack of algorithms and single algorithms from the set of best performed. The possible explanation is that better prediction is impossible for this data set regardless of the method used.

All in all, the following experiments were planned. The first is connected with the automatic stacking of algorithms by the creation of a complex decision tree. Moreover, the second experiment with the usage of the Auto ML procedure, which has the internal automatic process of creating new features, was planned.

4.2.8 Feature Reduction (Step 4)

The previous chapters presented the results of various algorithms used to predict the level (low, medium, high) of five personality traits according to Big 5. The predictions were based on a full set of created variables. Such extended models can be used when there are no time limits for the computation of the prediction. The aim of this dissertation is to build a way to predict personality in a very limited period of time as soon as possible after installing the application on a smartphone. Undoubtedly, the amount of processed data affects the time of calculating the predictions. The way to shorten the time is therefore to reduce the number of variables used in the calculations.The natural way is to limit the list of variables to only the most important ones. When designing a study, it is usually not assumed what variables will be significant in determining the traits level or class. Likewise, this study examines all available data in its initial stages. As the number of functions in a training dataset increases, the machine learning model becomes more complex, but also computationally expensive. To lower the complexity of the model, the goal may be to develop a trained machine learning model with the minimum number of required functions that can predict the personality profile yet with acceptable accuracy. The disadvantage would be both the oversimplification of the model and the loss of important information by trimming the features. In addition, there is also the danger of an overly developed, inefficient model based on too many variables without any gain in the form of quality improvement. Since one of the goals of this dissertation is to create a model that can be used during a fairly short installation time of the application on the phone, this optimisation is the way to achieve the goal.

Table 4.7 shows the feature reduction in percentages for each personality dimension in all data categories.

Table 4.7. Initial Feature Reduction based on Mutual Information value. The percentages are calculated on category base (rows).

	TOTAL	E	A	O	S	C	E	A	O	S	C
aplication list	113	57	50	42	67	60	50%	44%	37%	59%	53%
calls history	43	28	26	23	26	21	65%	60%	53%	60%	49%
contact list	30	22	12	15	21	17	73%	40%	50%	70%	57%
phone settings	75	37	40	45	35	36	49%	53%	60%	47%	48%
	261	144	128	125	149	134					

The first reduction was applied by cutting off all features with an MI value lower than

0.1 and supplementing these lists with features with the highest correlation coefficients with the predicted personality trait ('number of selected features column' in the table 4.11). Based on this set of variables, the predictive ability of the entire list of algorithms that can be used directly in the Android application was verified again. The summary for this verification is presented in the table 4.8.

Table 4.8. Verification of ML algorithms performance on initial shortened set of features. A is for Accuracy, PW is for Precision Weighted, and F1W is for F1 Score Weighted.

	E			A			C			S			O		
ALGORITHMS	A	PW	F1W	A	PW	F1W	A	PW	F1W	A	PW	F1W	A	PW	F1W
CART	.52	.57	.54	.50	.54	.52	.54	.59	.59	.51	.55	.53	.57	.62	.59
ADA	.57	.60	.58	.56	.55	.56	.61	.59	.60	.58	.57	.57	.62	.63	.62
ETC	**.71**	**.59**	**.62**	**.69**	**.57**	**.59**	**.73**	**.59**	**.62**	**.71**	**.57**	**.60**	**.75**	**.63**	**.66**
RF	**.69**	**.60**	**.62**	**.67**	**.55**	**.58**	**.73**	**.58**	**.63**	**.69**	**.56**	**.61**	**.74**	**.62**	**.66**
MLP	.62	.60	.61	.55	.55	.55	.62	.59	.60	.59	.56	.58	.62	.61	.62
Bernoulli NB	.35	.61	.38	.38	.57	.42	.31	.58	.34	.33	.59	.37	.34	.63	.40
Gaussian NB	.15	.46	.09	.16	.53	.10	.22	.56	.19	.20	.55	.18	.17	.54	.09
KNC	.31	.58	.35	.33	.54	.37	.30	.57	.33	.28	.58	.31	.31	.62	.36
Linear SVC	.36	.58	.39	.33	.56	.36	.34	.58	.38	.37	.58	.42	.35	.62	.41
NuSVC	.58	.59	.58	.51	.54	.56	.60	.59	.59	.56	.55	.56	.57	.63	.60
SVC	.55	.58	.56	.46	.54	.49	.57	.59	.58	.53	.56	.54	.53	.63	.57

Regardless of the personality trait, two algorithms that proved to be the best were RF and ETC.

Table 4.9. Comparison of the results for the two best algorithms (RF, ETC) from the training procedure (Hold Out with CV) and on the hold out test set not used for training the model.

	E			A			C			S			O		
ALGORITMS	A	PW	F1W	A	PW	F1W	A	PW	F1W	A	PW	F1W	A	PW	F1W
ETC - CV	.71	.59	.62	.69	.57	.59	.73	.59	.62	.71	.57	.60	.75	.63	.66
RF - CV	.69	.60	.62	.67	.55	.58	.73	.58	.63	.69	.56	.61	.74	.62	.66
ETC - Test	.71	.63	.61	.68	.49	.57	.74	.55	.63	.71	.52	.60	.73	.57	.64
RF - Test	.72	.70	.66	.70	.68	.61	.74	.55	.63	.67	.51	.58	.74	.57	.65

Consequently, only for these two algorithms (RF, ETC) the next RFECV procedure was iterated. RFECV procedure aims to find the smallest set of features that will be sufficient to compute the predictions with the qualitative parameters of the prediction not worse than the prediction on the full original feature set.

The *Scikit-Learn* library in Python software provides methods to simplify the model by further reducing the dimensions of the training dataset with the ability to control the impact of this reduction on the prediction accuracy of the machine learning model.

These analytical models are the Recursive Feature Elimination in the cross-validation process (REFCV) and Permutation Importance Evaluation (PI). In REFCV, the validity of a feature is calculated based on the selected estimator and the indicated number of features to be reduced (step), and in turn, each iteration of the algorithm discards the weakest features. The whole cross-validation procedure is the same as in step 3.

The validity of the permutation function is a model check technique that can be applied to any fitted estimator, and is especially useful for non-linear or opaque estimators. Permutation Importance is defined as a decrease in the model's score (e.g. Accuracy) when a single feature value is randomised. This procedure removes the existing relationship between the explanatory trait and the target. Then the decline in the model score indicates how much the entire model depends on the feature. Since the features considered to be of little importance for a bad model may be very important for a good model - this analysis only confirms the results of the RFECV obtained (Breiman, 2001).

Table 4.10 presents the complete list of features (with data category affiliation) used for the UISPP model creation. Categories like Application List or Phone Settings do not require User access consent. Features from the Contact List category and Calls History require separate consent for access from the user.

Table 4.10. The complete list of features used for creating UISPP with data category. Cat. is for Category. Abbreviations: AP is for application list, CH is for calls history, PS is for phone settings, CL is for contact list

Extraversion (35)	Cat.	Agreeableness (50)	Cat.	Conscientiousness (9)	Cat.	Stability (40)	Cat.	Openness (15)	Cat.
category_is_Books&Reference_ratio	AP	animator_duration_scale	PS	category_is_other_ratio	AP	all_apps	AP	category_is_Music&Audio_sum	AP
category_is_Business_ratio	AP	apps_from_GooglePlayStore	AP	system_apps	AP	available_size	PS	category_is_Other_ratio	AP
category_is_Communication_ratio	AP	available_size_sd	PS	contacts_mobile	CH	category_is_Business_ratio	AP	category_is_Productivity_ratio	AP
category_is_Education_ratio	AP	bluetooth_on	PS	type_is_missed_ratio	CH	category_is_Entertainment_ratio	AP	category_is_Tools_sum	AP
category_is_Entertainment_ratio	AP	category_is_Action_sum	AP	games_ratio	AP	category_is_Finance_ratio	AP	category_is_Travel&Local_ratio	AP
category_is_Entertainment_sum	AP	category_is_Adventure_sum	AP	category_is_Shopping_ratio	AP	category_is_Finance_sum	AP	category_is_Video Players_ratio	AP
category_is_Finance_ratio	AP	category_is_Books&Reference_ratio	AP	number_type_is_fixed_line	CL	category_is_Food&Drink_ratio	AP	contacts	CL
category_is_Music&Audio_ratio	AP	category_is_Books&Reference_sum	AP	category_is_Communication_ratio	AP	category_is_Health&Fitness_ratio	AP	contacts_mobile	CL
category_is_Productivity_sum	AP	category_is_Business_ratio	AP	duration_calls_mean	CH	category_is_Maps&Navigation_ratio	AP	contacts_type_mobile	CL
category_is_Puzzle_ratio	AP	category_is_Business_sum	AP			category_is_Music&Audio_ratio	AP	duration_calls_max	CH
category_is_Tools_sum	AP	category_is_Card_sum	AP			category_is_Other_ratio	AP	duration_calls_std	CH
contact_one_month	CL	category_is_Comics_ratio	AP			category_is_Other_sum	AP	duration_calls_sum	CH
contact_three_months	CL	category_is_Communication_ratio	AP			category_is_Photography_ratio	AP	type_is_INCOMING	CH
contact_two_weeks	CL	category_is_Communication_sum	AP			category_is_Productivity_sum	AP	user_apps	AP
contacts	CL	category_is_House & Home_sum	AP			category_is_Social_ratio	AP	weekend_ratio	AP
contacts_family	CL	category_is_Libraries&Demo_ratio	AP			category_is_Social_sum	AP		
contacts_fixed	CL	category_is_Libraries&Demo_sum	AP			category_is_Tools_ratio	AP		
contacts_mobile	CL	category_is_Medical_ratio	AP			category_is_Tools_sum	AP		
contacts_type_mobile	CL	category_is_Personalization_sum	AP			category_is_Travel&Local_ratio	AP		
contacts_with_duplicated_numbers	CL	category_is_Productivity_sum	AP			category_is_Video Players_ratio	AP		
duration_calls_max	CH	category_is_Social_ratio	AP			contact_one_week	CL		
duration_calls_sum	CH	category_is_Social_sum	AP			contact_three_months	CL		
games_ratio	AP	category_is_Strategy_sum	AP			contacts	CL		
number_type_is_mobile	CL	category_is_Tools_ratio	AP			contacts_family	CL		
number_type_is_mobile_ratio	CL	category_is_Tools_sum	AP			contacts_fixed	CL		
number_type_is_short_ratio	CL	category_is_Travel&Local_ratio	AP			contacts_mobile	CL		
screen_brightness	PS	category_is_VideoPlayers_sum	AP			contacts_type_mobile	CL		
system_apps	PS	contacts	CL			contacts_unknown	CL		
total_bytes	PS	contacts_family	CL			count_calls	CH		
type_is_incoming_ratio	CH	contacts_foreign	CL			density_dpi	PS		
type_is_missed_ratio	CH	contacts_mobile	CL			duration_calls_median	CH		
type_is_outgoing	CH	contacts_photo	CL			duration_calls_std	CH		
type_is_outgoing_ratio	CH	contacts_type_mobile	CL			free_size	PS		
type_is_rejected	CH	dtmf_tone_when_dialing	PS			games_ratio	AL		
work_period_ratio	AP	font_scale	PS			games_sum	AL		
		free_size_sd	PS			number_type_is_fixed_line_ratio	CL		
		mean_day	AP			number_type_is_mobile	CL		
		number_type_is_fixed_line_ratio	CL			number_type_is_mobile_ratio	CL		
		number_type_is_foreigen	CL			type_is_incoming	CH		
		number_type_is_foreigen_ratio	CL			user_apps	AP		
		number_type_is_toll_free_ratio	CL						
		rotation_enabled	PS						
		total_bytes	PS						
		total_bytes_sd	PS						
		total_size_sd	PS						
		tts_default_rate	PS						
		type_is_missed_ratio	CH						
		type_is_unknown_ratio	CH						
		version_sdk	PS						
		weekend_ratio	AP						

Source: own work

Table 4.11. Results of iterative feature reduction procedure with final feature numbers used for prediction.

Big 5 dimension	Number of features based on MI	Number of selected features	Steps in reduction on iteration	Final number of features used for creating UISPP
E	144	97	RFECV RF: 66, 52, 41, 40 PI RF: 40, 37, 36, 35	35
A	128	60	RFECV RF: 50 RFECV ETC: 40, 11	50
C	134	109	RFECV RF: 19, 9	9
S	149	98	RFECV RF: 31, 29, 28, 27 RFECV ETC: 49, 40	40
O	125	75	RFECV RF: 14 RFECV ETC: 15	15

The final step in building the final solution is to look for the best values for hyperparameters, which cannot be learned directly in training and looking for estimators (L. Li et al., 2017).

Table 4.12. Final Random Trees Classifier Parameters used for Predicting Each Personality Dimension Class

	E	A	C	S	O
Classifier	RF	RF	RF	ET	RF
random_state=	0	0	0	0	0
max_depth=	30	22	25	25	30
max_features=	1	7	3	4	1
min_samples_leaf=	1	1	4	2	1
n_estimators=	95	400	350	400	400

In *Scikit-Learn*, they are given as arguments to the constructor of the estimator classes. Using the trial and error method, the given parameters are optimized to obtain the highest possible measures of the model quality (Accuracy, Precision, F1 score). Table 4.12 presents the final settings for the parameters describing the constructed random trees. Parameterisation concerns the limit for the maximum length of random trees to be built, the number of features and leaves in the trees, and the total number of built estimators.

4.2.9 Final UISPP Model

In the previous section, the results of personality recognition are discussed using the various machine learning techniques available. Based on the examination results, it can be concluded that no single method works best for all Big 5 dimensions. Also, there is no one-way consistency in selecting the best method, which gives the best Precision and Accuracy in a repeatable way. It should also be noted that considering project objectives, the measure of Precision seems to be more critical. In the initial model building stages, both the predictive power of the prepared lists of features and the effectiveness of the techniques of building the model were tested. These steps are needed to test the predictive power of data. However, due to the defined goals and criteria that the built model of determining personality must meet, it is necessary to finally reduce the computational algorithm by limiting the features and parameters of the model. In this last step, the procedure is not to look for the best quality indicators (Accuracy, Precision), but rather to reduce the algorithm's complexity while maintaining the quality parameters. It is carried out by manually adjusting the limits for tree expansion and looking for the best quality configuration personalisation mechanism.

Table 4.13. Comparison of Algorithms: Accuracy (A), Precision weighted (PW) and F1 Score weighted (F1W) for unlimited tree parameters (Test1) and limited number of trees up to 100 (TEST2)

	Optimal UISPP model (N=100)			Final UISPP model (N=100)		
	A	PW	F1W	A	PW	F1W
EXTRAVERSION RF (35 features)	.76	.74	.73	.76	.74	.73
AGREEABLENESS RF (50 features)	.72	.76	.63	.71	.62	.61
CONSCIENTIOUSNESS RF (9 features)	.69	.57	.62	.70	.61	.64
STABILITY ETC (40 features)	.69	.58	.62	.71	.65	.63
OPENNESS RF (15 features)	.73	.64	.68	.74	.65	.69

In the table 4.13 final parameters of the UISPP model are presented for each personality trait. It summarises the optimal final model created on a small number of features, from 9 to 50, depending on the dimension (given in brackets after dimension name). It presents results from the optimal UISPP model created according to parameters from table 4.12. Unfortunately, after the first attempts to implement the model directly to the android, it turned out that it is too extensive. The current capabilities of the android system did not allow such a large number of trees to be made. For this reason, it is decided to reduce the number of trees to 100 per feature - this number of trees did not cause problems with the code. The results in part of the final UISPP model relate to a model based on the same variables but for the number of trees reduced to 100 per feature.

Table 4.14. Final UISPP model parameters (Accuracy, Precision, Recall) in comparison to the best in class model based on all features (not reduced)

		E	C	A	S	O
Best in class with not reduced number of features	Number of features based on MI	144	128	134	149	125
	Method	AML	AML	SC	AML	ML
	Accuracy	.77	.75	.71	.72	.74
	Precision	.72	.68	.67	.67	.59
	F1 score	.69	.65	.60	.62	.65
Final UISPP model	Number of features based on RFECV	35	50	9	40	15
	Method	RF	RF	RF	ETC	RF
	Accuracy	.76	.69	.72	.69	.73
	Precision	**.74**	.57	**.76**	.58	**.64**
	F1 score	**.73**	.62	**.63**	.62	**.68**

MI-Mutual Information, AML-Auto ML, SC-Stacking Classifier, RF-Random Forrest Classifier, ETC-Extra Tree Classifier

Finally in the table 4.14, parameters of the final UISPP model are compared to the best achieved models from the first stages of model creation. It can be treated as a kind of internal validation of the final model. The best in class models is performed on the mot reduced number of features without reduced number of trees. Comparing the final UISPP model results with this best achieved model we can see that in some cases the parameters are even slightly higher. It takes place in case of Extraversion, Agreeableness and Openness.

4.2.10 Implementing UISPP Model in App

Currently, a commonly used solution is the implementation of intelligent algorithms based on data using data collection in the cloud. However, in the case of cloud-dependent systems, disregarding the issue of privacy and necessity of sharing private data, there is still a technical problem related to the internet connection. Any solution based on cloud computing is exposed to the risk of data unavailability when there is no Internet access (Nadeski, 2019). In the field of Internet communications, new solutions are constantly emerging to improve the stability of the Internet connection, traffic, data transmission speed and latency. Moreover, multiple computations are needed to understand data behaviour and to generate data representations used in inference. In order to transfer the model from analytical software (Sci-learn Python) to an application processing data on the device, there is a need to use transpilers, i.e. exporters converting code with a high level of complexity and computational needs to code with a low level of complexity (Branco et al., 2019) .

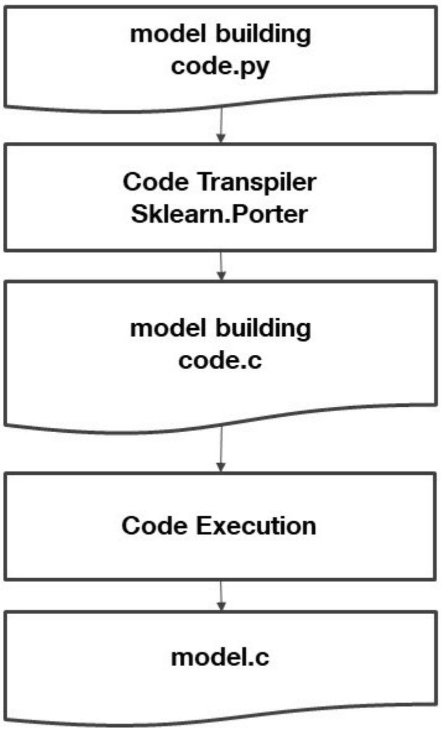


Figure 4.18. Transpiler's workflow. Based on the publication of (Branco et al., 2019), page 9

(Source: own work)

The code transpilation method (Fig 4.18) was used to build the second artifact. Code transpiling means that source code written in Python is converted into source code written in Java (represented by model.c on Branco et all scheme). The entire process of calculating the model will take place on the user's end device (smartphone). There are some transpilers available (Branco et al., 2019). Among those applicable for Java the Sklearn-porter[1] was chosen which is in the package for Python. Sklearn transpiler allows to transpile models trained with Scikit-Learn to other programming languages and Java is one of them. It's executed by converting the phyton code into model building code.c including the data standardisation process, executing it, and creating the model.c for calculating all transpilled decision trees (on Branco et all scheme

[1]Copyright 2016-2019, Dariusz Morawiec

c represents programming language C). There are some methods to transpile trained models from high-level languages to Java. The transpilers have also some limitations, the major one is that they do not make any code optimisation. Therefore, any code optimisation must be done by the developer.

4.3 Personality Based Automatic Personalisation of App Method

The created personalisation method is based on the user's personality profile, calculated automatically based on available data as soon as the service is installed. Knowing the user's personality profile has many advantages, such as fine-tuning the initial interface configuration and feature presentation (better first impression, satisfaction and user experience). It will also enable individualised support in the adaptation process and allow the simulation of the best-matched personality of a virtual agent. Thanks to the customer's individualised experience, the likelihood of service rejection will decrease, and satisfaction and loyalty will increase. This profile can also be used to implement an adaptive strategy in more advanced technologies to reduce cognitive (communication) barriers and technological fears. It will also allow the simulation of the personality best suited to the user when the service uses intelligent bots.

The proposed concept of personalisation (PAPA) consists of implementing the model of calculating the user's personality profile in the application itself. The user's personality profile is calculated automatically. The diagram (Fig 4.19) shows the process flow. The user installs an application on the smartphone with the UISPP model implemented (any service). The application counts the set of statistics needed to calculate the profile. During the standard adaptation and authorisation process, the application counts the statistics from the data on the phone, but only those needed to calculate the profile. In addition, with the user's consent, it automatically adjusts the appearance, functionality and communication to his/her profile. The profile remains secret and private, available only to the service and the user's end device. The application has the same functionalities for everyone, but thanks to the knowledge of the profile, it differentiates the way the service is delivered to the user. In addition, the application adapts to the user's capabilities; for example, people with low conscientiousness will get more reminders of what to do. The reinforcements (feedback) from the service are also tailored to the needs of each user: for extraverts social issues are underlined, and for people with high conscientiousness, the information received is focused on goals and the level of task performance.

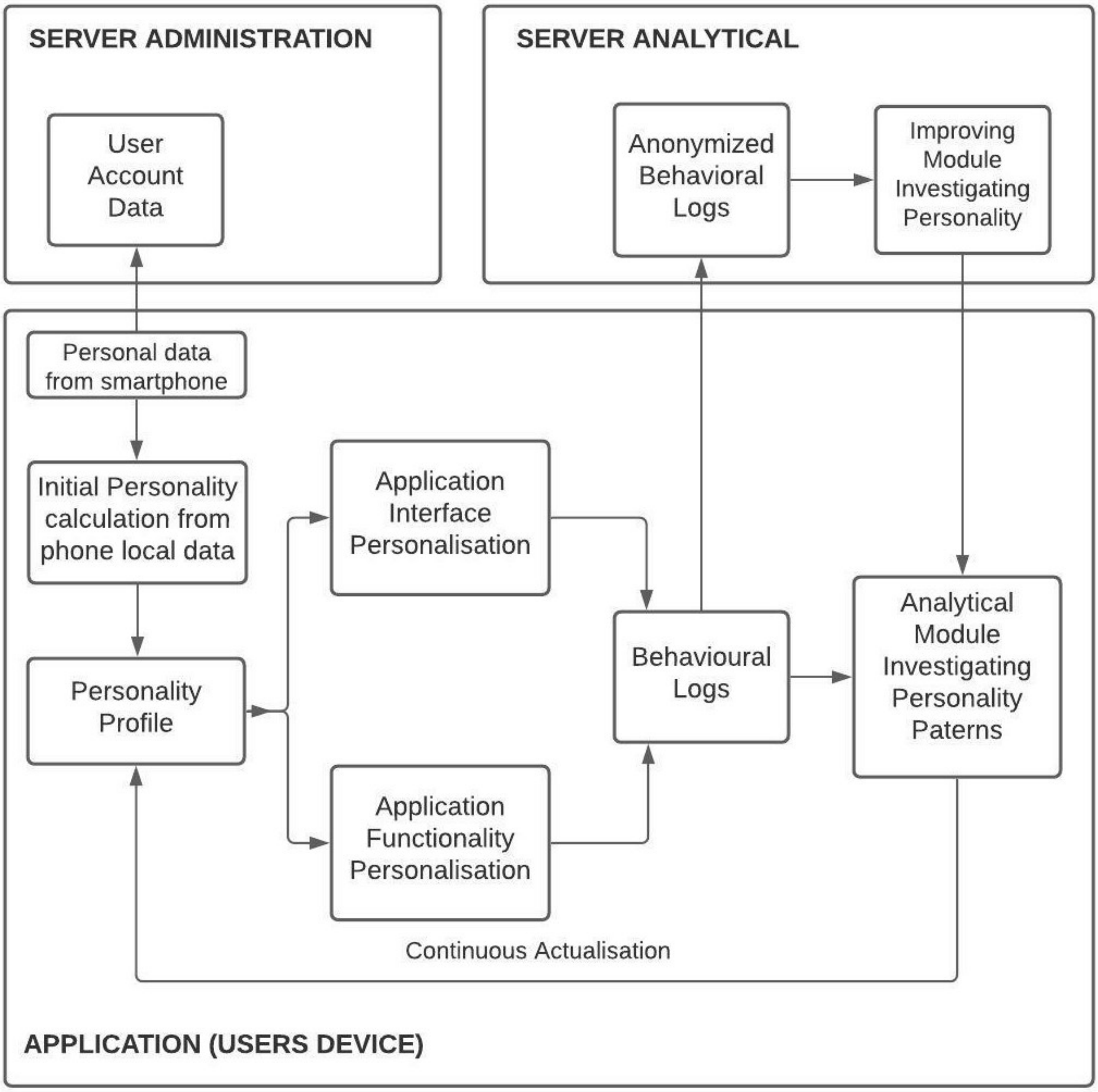


Figure 4.19. Architecture for Personalisation components inside the app

(Source: own work)

Adapting the service to different profiles requires additional research, which is developed to differentiate specific functionalities of the service. Additionally, the application analyzes continuously when the user's behaviour (choices) during the application's operation correspond to the calculated profile when using the application service. Based on these analyses, the application autonomously decides whether to continue profiling based on its initial profile or modify the profile based on the history of service use. It seems that there is no need to have a detailed personality profile (e.g. on a sten scale). It is enough to classify the user into the following classes: low, medium and high according to the level of a given feature. A division into three classes was used in the present study, defined by the boundaries between sten 3 - 4 and 7 - 8. Moreover, because various personalised elements on the service are matched independently to the level of individual personality dimensions, potentially, the entire profile does not necessarily have to be counted and used in each service. UISPP models are created independently for each dimension and can also be used independently. While adjusting the interface graphics, the only important thing is to differentiate between the dimensions of Extraversion and Openness. In this case, it is not necessary to count the entire profile (five dimensions). The separation of profile dimensions makes it possible to eliminate redundant information from processing (method ergonomics) and generating user knowledge (GDPR).

Accounting the profile locally on the user's end device (smartphone) eliminates the need to transfer data from the phone and the profile to the databases in the service's back-end. The profile is available only for use by the profile owner (end device). In addition, the profile is based on a small amount of data available at the time of service installation, thanks to which the automatic profiling of the service is possible from the first use. In addition, this model meets the assumed conditions, i.e. it is calculated based on data found on the phone during the installation of any application. The logic of automatic personalisation of the application based on the BIG profile 5 on the user's end device is shown in the Figure 4.19. All the necessary steps of the personalisation process, such as calculating the personality profile (UISPP model), the automatic adjustment of the user interface, the automatic adjustment of the application (functionalities), are carried out on the end device. The user grants access to the data on the app's phone for processing, but no personal data is sent to the app's back-end servers for verification, administration, or analytics management. Some anonymous statistics from the behavioural logs are sent when the application is updated for analytical purposes. This cycle allows the updating of standardised metrics for assessing the personality patterns required to monitor the accuracy of the personality profile. In this way, the means and standard deviations used in standardisation procedures can be updated.

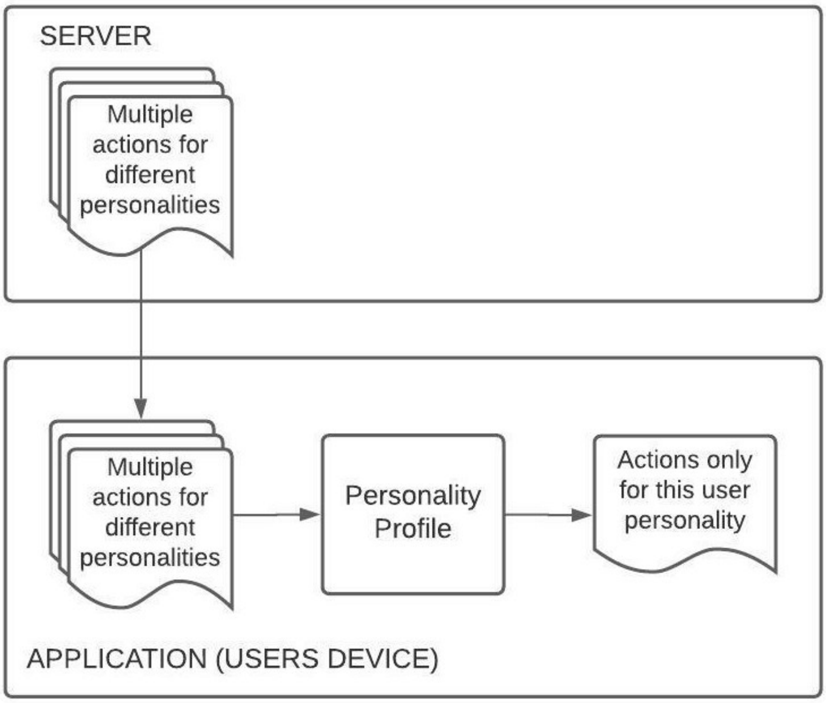


Figure 4.20. Architecture for additional Personalised Action initiated by the App Vendor (Source: own work)

The Figure 4.20 also shows the simplified concept of using a fully protected profile for external actions initiated by the application vendor. The supplier will prepare activities dedicated to particular groups with a specific level of functionality in the Big 5 profile. The application would receive the necessary changes in the code through the update process, but the application on the end device would decide whether the action should be launched for the user. This process would follow this scheme only for users who do not consent to share their data and supplier profile information. In this way, the provider would not have to limit its marketing activities to users who provide data and information about their behaviour. The mechanism would enable individualisation of personalisation despite opposition to sharing personal data. This mechanism theoretically enables the resale of marketing activities to third parties without providing customer data.

The evaluation process of proposed method of personalisation is presented in the next Chapter no.5.

4.4 Limitations

The UISPP model and PAPA method presented in this chapter have limitations. Some of them are related to the assumed business requirements, such as the automatic operation of the method only in the environment available on the end device. In the case of the model, the primary limitation is that the short time during installing the application forces time limits for the duration of the calculations. Therefore, in the final version of the UISPP model, the number of processed data had to be reduced despite the consequences of worse prediction accuracy parameters. Another limitation is related to the characteristics of the available data in the phone. Based on the literature review and conducted experiments, data that is more closely related to user decisions and derives from digital behaviour (e.g. written texts or data from application logs) is a more prominent source of data for personality prediction. The conducted experiments related to the construction of the model have shown that the quality of the model parameters depends more on the set of data used for prediction than on the machine learning methods used. Without extending the list of variables used to build the model, quality improvement will not be possible. An additional limitation is the lack of existing research on the definition points of the classes used (high, medium-low level of the feature) and the differentiation of behaviours. It was not the objective of these studies, but extending the research to this aspect could possibly increase the effectiveness of prediction. The differentiation of behaviour in groups is a direct factor influencing the prediction. The smaller the differentiation between groups, the less chance of the model's universality because of high behaviour interference . Another limitation related to the construction of the model is the time stability and cultural universality for the relationships between data, behaviour and the personality profile. This problem was partially addressed in the mechanism for verifying means and deviations from data used for prediction (PAPA methods). However, the UISPP update mechanism also needs to be refined. These mentioned limitations related to the prediction of the personality profile according to the UISPP model are also a limitation for the PAPA method itself. For this reason, the profile has been defined as initial in the method, and the auto-updating method has been designed based on data from the service. Because the use of the PAPA method itself requires additional research related to the design of personalisation in a specific service, this element could not be evaluated in the final service. The lack of availability of a service that uses the PAPA method is a limitation for its final evaluation. However, full implementation will help make the idea real without changing the overall concept, the main benefit of which is to leave the user's data in the realm of his private data not shared with service providers. One should also consider the simultaneous coexistence of the proposed model with the model based on sharing data with the service provider(to which a large proportion of customers agree), as it is currently the case .

4.5 Summary

The chapter presented the path of the UISPP model development (Artefact 1) and the general concept of the PAPA method (Artefact 2). The purpose of the first artefact is to classify clients in terms of their personality profiles. The personality profile reflects their needs and expectations about services and the approaches to communication. The second is a method for automatic personalisation of the application (or any service) based on the knowledge about the user constructed on the profile while maintaining complete customer data privacy. The concept of automatic personalisation of the application found on the data in the smartphone, keeping both the data and the user profile confidential, was defined based on the review of existing research results. The analysis of business needs, the gap in the existing knowledge base (Chapter 2), and customer expectations regarding using their data and automatic personalisation of services in preliminary research (chapter 3) preceded the creation of the final solution. The original solution creation started with the model building based on machine learning techniques. The model for classifying customers established by the personality profile calculation was created using data from 2,666 mobile phone users. From the initial 209 recorded statistics from the phone, there are generated 477 variables describing human behaviour. Step-by-step stages are outlined to find the method that best suits these data categories and best predicts the user's personality. The iterative procedure in the first stage was to create a model with the best parameters (Accuracy, Precision, F1 score). The next stage aims to reduce the model to a version that could be implemented on an end device with limited computing power. The final UISPP model obtained parameters exceeding most of the assumed baselines, including similar models from the literature review. The ranges for the 5 features are: Precision: 0.70-0.76, Accuracy: 0.61-0.76

and F1: 0.69-0.71. Next, the method of implementing the prediction mechanism directly in the mobile application is described. Finally, the logical architecture of using the personality profile for personalisation was presented. It will be used in building an application prepared for UX / UI tests, taking into account the personalisation of the application. The proposed solution answers three research questions defined in chapter 1 of this dissertation. The research has shown that it is possible to predict the user's personality based on static data from the phone, implement such a predictive model in the phone, and use it for an automated application personalisation mechanism.

Chapter 5

Evaluation

5.1 Evaluation Methodology

According to the guidelines contained in the Design Science methodology proposed by (A. Hevner and Chateerjee, 2012), each performed study is evaluated. The previous chapter describes the artefact creation process, and the current one describes the evaluation method and results. First, the evaluation methodology used in the dissertation will be described. Then, the following two sections summarise the evaluation for the user's smartphone initial personality profile (UISPP) model (Artefact 1) and the PAPA method in the test smartphone application (Artefact 2). The evaluation is done on data from the research conducted for this dissertation. Finally, the chapter presents the evaluation's conclusions and a summary.

5.1.1 Framework for Evaluation in Design Science Research

The general assessment methodology presented in this chapter is in line with the Framework for Evaluation in Design Science Research (FEDS) described by (Venable et al., 2016). According to the FEDS guidelines, the evaluation process consists of four main steps:

1. Defining the evaluation objectives, i.e. their precise explanation, emphasis on the elements such as rigour, ethics, efficiency, uncertainty, and risk reduction.

2. Choosing an evaluation strategy tailored to the goal defined in the first step.

3. Determining the properties that will be assessed. This step selects a detailed set

of functions, goals, and requirements for evaluating the artefact.

4. Designing the individual activities needed to implement the evaluation and carrying out this evaluation for each defined artefact.

According to the FEDS classification proposed by Venable et al. (2016), the evaluation applied to the research presented in the dissertation is an ex-post evaluation, in line with the risk and effectiveness strategy envisaged for the created technical implementations whose recipient is the user. 'Ex-post' and 'summing up' refer to the assessments carried out after the research and at the end of the development process after the execution and implementation of artefacts. Functionally, the purpose of 'summative' evaluations is to assess the degree to which the obtained artefact meets users' expectations and to what extent it is effective. The FEDS evaluation framework distinguishes between two evaluation paradigms: artificial evaluation and naturalistic evaluation (Venable et al., 2016). Artificial evaluation includes laboratory experiments, simulations, criteria analysis, theoretical arguments and mathematical proofs. On the other hand, artificial evaluation is dominant in the scientific and rational paradigm and leads to an artificial evaluation of the benefits of greater scientific credibility through repeatability and falsifiability. The second type of evaluation is naturalistic evaluation, which aims to test the performance of a solution in a real environment within the organisation. When assessing a real environment, e.g. in business (real people, real systems), the complexity and volition of the human factor and the organisation's practices are considered. Naturalistic evaluation methods are typically case studies, field studies, field experiments, surveys, ethnography, phenomenology, hermeneutic methods, and action research. Both an artificial (criterion analysis, simulation) and naturalistic (test in action and customer evaluation) evaluation method are provided for the current artefacts. The separate artefact evaluation process and criteria were designed for each artefact.

User research (naturalistic evaluation method) was carried out on 170 volunteers recruited by social media using a specially designed application. After being installed on a smartphone, an application standardises the data needed to calculate the UISPP, converts the model, and shows the profile results on the screen. Additionally, in the application, the user voluntarily answers four questions: age (years), gender (female, male) and two additional evaluation questions on the 7-point, balanced, LIKERT scale:

Question 1: *'To what extent does the presented personality profile fits you?'*

Question 2: *'Would you allow the application on your phone to automatically adapt its operation to such a profile?'*

The Likert scale is named after Rensis Likert, who developed it in 1932. It is one of the most widely used and well-studied techniques used in questionnaires. It is used to obtain knowledge about the degree of acceptance of phenomena, views, processes, features and attitudes.(Magnusson, 1967)

These two standard questions were intended to quantify the acceptance level for an automatically computed personality profile and for the very concept of using it for personalisation. It seemed crucial to separate the response to the profile itself from the personalisation question, which was the primary goal of this measurement. Answering both questions was not obligatory in the technical test. 129 out of 170 participants answered the questions.

The application sent the model calculation results and answers to the questions directly to the server, which made it possible to check both the times in which the model counts on each device as well as collecting information about how many people with special needs (level, any feature in the low or high class) can be identified using the model.

5.1.2 Evaluation of UISPP model

Based on the FEDS, the following evaluation criteria for the UISPP model were defined:

(C.1) Accuracy, Precision and F1score of the UISPP model.

(C.2) Calculation performance of the UISPP Model under the conditions required by the business (time of computing the personality profile).

(C.3) User acceptance of the automatic method of calculating personality based on smartphone data.

In this study, Accuracy refers to the percentage of correctly predicted values of the Big Five personality dimensions determined by the questionnaire measurement (described in Section 3.4 and Appendix B). Accuracy is calculated for a three-class distribution (Low, Medium, High) as described in Section 4.3.4. Method and mathematical formula of Accuracy and other metrics used for evaluation are outlined in Section 2.3.2. The evaluation was carried out separately for each of the five personality dimensions because the final UISPP model consists of five models, one for each personality dimension.

First, a UISPP model was built as described for the assessment presented in Section 4.3. Precision, Accuracy, and F1score were then calculated to evaluate the quality of the model created. Ultimately, the model's validity was calculated on an independently drawn sample of stratified randomised data from the study described in Section 3.5. according to the envisaged model building procedure (Section 4.3.6). The data used for evaluation are held out before building the model and was not used in the machine learning process, i.e. the learning and testing cycle in which the model is created.

Finally, the UISPP model has been empirically tested during technical tests. The technical feasibility of the model implemented in the Android application was assessed. The purpose of the technical tests was primarily to check the time range and processor load. Such a measurement is pivotal for assessment, thanks to a model constructed in this way, it is possible to calculate classes for personality traits in a sufficiently short time. The speed of model calculation, in turn, is a prerequisite for using the PAPA personalisation method from the beginning of using an electronic service (e.g. a mobile phone). The results of the evaluation of the three-stage model are presented in Section 5.2. In addition, the participants of the UX tests accompanying the PAPA model evaluation also assessed the compliance of the profile calculated from the data thanks to the UISPP model. As a result, the test was supplemented with a rating from a real user.

5.1.3 Evaluation of PAPA method

The criteria for PAPA method evaluation were defined during the research process and tested in a prototype built especially for this purpose (according to FEDS). The selected evaluation criteria relate to:

(C.1) Feasibility of the method in real environments (mobile application prototype)

(C.2) Usability of the method to function despite not sharing user data and profiles derived from this data with the service provider.

(C.3) User acceptance of the model personalising the service based on a personality profile calculated from smartphone data.

As mentioned earlier, the PAPA method takes the results of the UISPP model as input, i.e. the user profile. The PAPA method is based on the automatic adjustment of service parameters (e.g. interface, functionality, communication with the user, frequency of communication with the user, principles of operation of referral systems).

At this stage, without additional research related to the execution of personalising a specific service, its functionality, content recommendation, interface appearance, and communication method, it is not possible to fully assess the effectiveness of the personalisation created in this way. There is not enough knowledge base to make such decisions.

Creating the final shape of personalisation to the customer requires a lot of additional independent information and existing service delivered with the smartphone application. Furthermore, the response is strongly contextual, as exemplified by research provided by (Matthews, Lin, et al., 2020), (Matthews, Hancock, et al., 2020) and (Szalma and Taylor, 2011). According to the conducted research, the impact of personalisation on the basis of the user's personality cannot be considered or studied without taking into account important factors influencing behavioral intentions, such as the usability of the application / service for the user, the real need to use the system or general attitudes towards technology (Davis et al., 1989). People first decide to choose a specific service or software that serves a specific purpose, and only after this decision, other factors that can be influenced by personalisation are important (Svendsen et al., 2013). This dissertation aims to test the technical feasibility and usability of the Service Personalisation Concept (PAPA) based on the Personality Profile (UISPP). A study on potential users was prepared, and an assessment was done only for the concept of personalisation. The feasibility assessment and overall user acceptance of the PAPA method are presented in chapter 5.3.

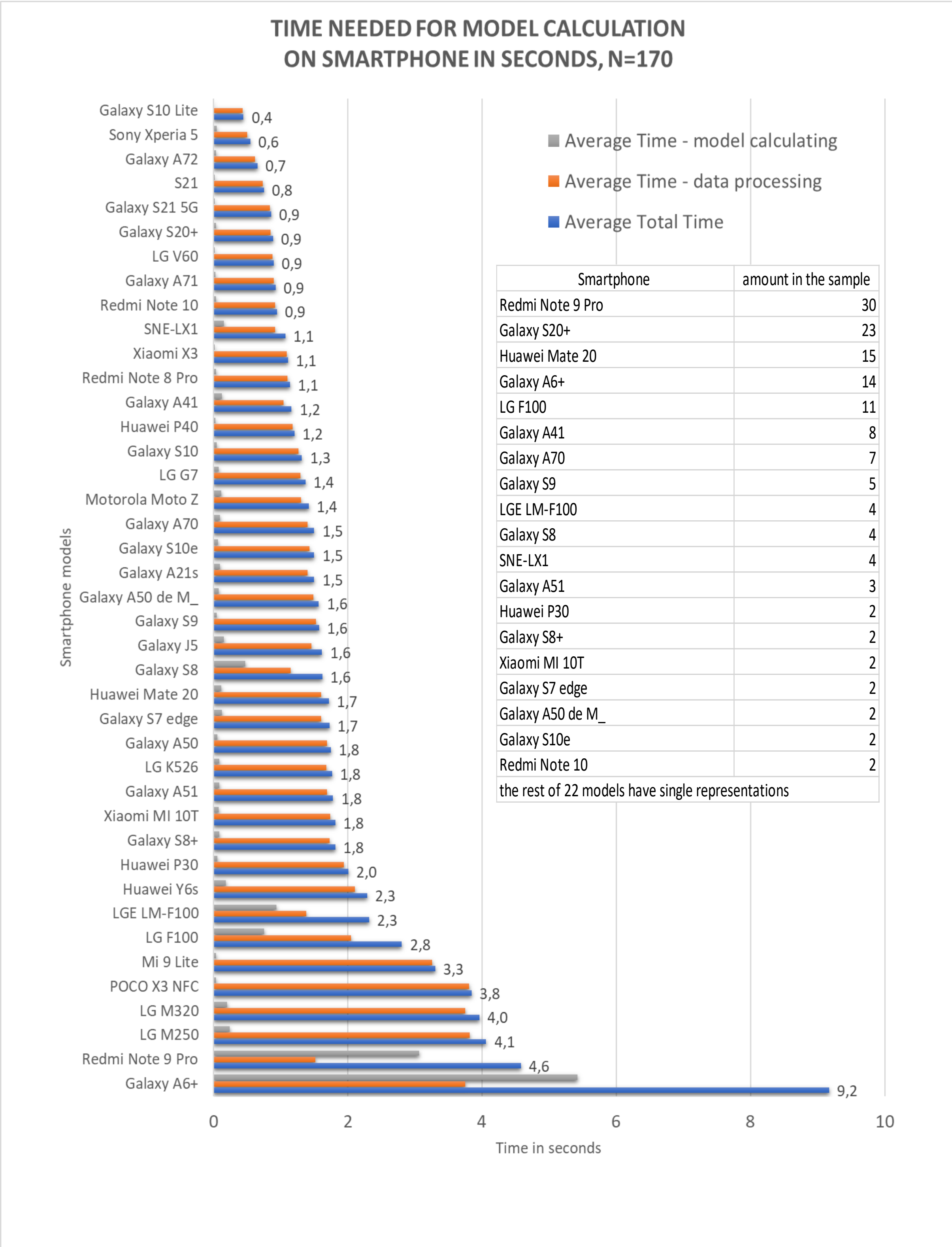
5.2 User's Initial Smartphone Personality Profile Model Evaluation

The description of the model creation procedure is presented in the chapter. The trustworthy parameters (Precision, Accuracy, F1 score) of the UISPP Model has been validated based on the conducted experiments. Each time, the evaluation was carried out on the hold out data set (not participating in the model's training) - 100 people were randomly selected for each test. Table 5.1 shows both test results and baseline levels for the adopted measures of the model assessment (Precision (weighted), Accuracy (weighted), F1 score (weighted). Because the model predicts three levels of personality traits naturally imbalanced (issue described in chapter 4.3.4), the measures are weighted by the class structure). Accuracy as a measure of model goodness is not perceived as the best measure by ((Fiorentini and Losa, 2020) (J. Zhang et al., 2020) for imbalanced classes. Unfortunately, as presented in the review in Section 2.3.3, Accuracy is the most popular measure of comparing the effectiveness of the models (in general). It is a consequence that models are usually based on binary classification (the population is divided into two halves according to the median).

Regarding the random benchmark (described in Section 4.3.4), the final model is significantly better than randomly assigning a class. In the case of the personalisation baseline, i.e. informing about the relation between the use of the model and its absence (then theoretically only the middle class receives the adjusted product), the final UISPP model has better Precision and a better F1 score measure for all features except Conscientiousness where the model has a lower but comparable F1 measure. In the case of Accuracy, the final model has higher measures only for the Extraversion feature. Contrary, the reference values for similar but not the same models from the literature review do not exceed this value of the personalisation baseline. As mentioned above, Accuracy is not the best measure to judge the predictions of unbalanced classes. In such a situation, Accuracy can only be high by meeting one major class, and this is the case for a computed baseline where we consider all to be middle class (4.5). However, in the context of using the model, the high Accuracy achieved by the mean class is not sufficient to evaluate the model. Because the model is dedicated to personalisation, the most required is discovering the highest possible percentage of those with any personality trait in the low or high class. Therefore, Precision and F1 score (the geometric mean of Precision and Recall) better evaluate the non-binary, imbalanced classes model. Taking these remarks into account, it can be concluded that the achieved results are satisfactory. The table also includes the evaluation results of previous large models. It should be noted that the final UISPP model, although on a smaller amount of data and less extensive trees, also for all traits, is better than the best of the large models for 4 of the five personality traits.

Table 5.1. Comparison of All Model Built to Assumed Benchmarks

	Extraversion			Conscientiousness			Agreeableness			Stability			Openness		
	P	A	F1	P	A	F1	P	A	F1	P	A	F1	P	A	F1
Baseline random	0.56	0.34	0.39	0.60	0.35	0.41	0.54	0.34	0.38	0.57	0.35	0.40	0.63	0.35	0.41
Baseline personalisation	0.52	0.72	0.61	0.55	0.74	0.63	0.49	0.70	0.58	0.52	0.72	0.60	0.58	0.76	0.66
Baseline SOTA								A: 0.61- 0.69							
S1. Holdout	0.66	0.7	0.68	0.55	0.74	0.63	0.62	0.69	0.60	0.52	0.72	0.60	0.67	0.76	0.68
S1. Holdout with Cross Validation	0.65	0.73	0.65	0.52	0.72	0.60	0.68	0.71	0.60	0.52	0.72	0.60	0.59	0.77	0.67
S2. Soft Voting	0.65	0.74	0.65	0.55	0.74	0.63	0.58	0.70	0.59	0.52	0.72	0.60	0.59	0.76	0.66
S2. Hard Voting	0.65	0.74	0.65	0.55	0.73	0.62	0.67	0.71	0.60	0.52	0.72	0.6	0.59	0.76	0.66
S3. Stacking Classifier	0.70	0.74	0.64	0.55	0.74	0.63	0.67	0.71	0.60	0.52	0.72	0.60	0.59	0.76	0.66
S.4. Auto ML	0.72	0.77	0.69	0.68	0.75	0.65	0.49	0.69	0.57	0.67	0.72	0.62	0.59	0.74	0.65
UISPP (optimal trees no.)	0.76	0.74	0.73	0.72	0.76	0.63	0.69	0.57	0.62	0.69	0.58	0.62	0.73	0.64	0.68
Final UISPP (trees no. <100)	0.76	0.74	0.73	0.71	0.62	0.61	0.70	0.61	0.64	0.71	0.65	0.63	0.74	0.65	0.69
UISPP vs baseline 1	0.20	0.40	0.34	0.11	0.27	0.2	0.16	0.27	0.26	0.14	0.3	0.23	0.11	0.3	0.28
UISPP vs baseline 2	0.24	0.02	0.12	0.16	-0.12	-0.02	0.21	-0.09	0.06	0.19	-0.07	0.03	0.16	-0.11	0.03
UISPP model vs baseline 3 low		0.13			0.01			0			0.04			0.04	
UISPP vs baseline 3 high		0.05			-0.07			-0.08			-0.04			-0.04	
UISPP vs the best created	0.04	-0.03	0.04	0.03	-0.13	-0.04	0.02	-0.1	0.04	0.04	-0.07	0.01	0.15	-0.12	0.02



Smartphone	amount in the sample
Redmi Note 9 Pro	30
Galaxy S20+	23
Huawei Mate 20	15
Galaxy A6+	14
LG F100	11
Galaxy A41	8
Galaxy A70	7
Galaxy S9	5
LGE LM-F100	4
Galaxy S8	4
SNE-LX1	4
Galaxy A51	3
Huawei P30	2
Galaxy S8+	2
Xiaomi MI 10T	2
Galaxy S7 edge	2
Galaxy A50 de M_	2
Galaxy S10e	2
Redmi Note 10	2
the rest of 22 models have single representations	

Figure 5.1. Model calculation time on smartphone - distribution from tests with specification of smartphone model name. Separate times for: total time of calculation, time for data processing and time for calculating model. Data from technical test, N=170

(Source: own work)

The exception is the model for Conscientiousness, which has higher Precision but a lower F1score (similar to the comparison with the personalisation benchmark). In summary, concerning external research benchmarks, theoretical benchmarks, and internal benchmarks, the final models for UISPP features are comparable or even better accurate for at least four characteristics: extraversion, agreeableness, stability and openness. The UISPP model for the Conscientiousness trait is not better than the baseline for all measures, but it cannot be considered significantly worse either. From the point of view of the business use of models, such conditions are acceptable because individual features of Big 5 can be used separately. The literature review on the use of personality in the personalisation of electronic services shows that two features seem to be pivotal for personalisation: extraversion and openness. In addition, the literature review did not find comparable models based on smartphone data, static data not derived from observations over a more extended period. According to the literature review, it is the first such model to be counted on the user's end device without transferring its data to external servers. The experiments reported in the previous chapter prove that despite using a small amount of user data (without observation and an extensive history of behaviour), the created algorithms have comparable quality indicators as other similar algorithms determining the personality profile. However, this dissertation's main goal was to create the method of determining the user's personality based on data from a smartphone with the efficiency that allows it to be used during the installation of the service on a smartphone. The average times in which the model counts were calculated were reported to check this condition. It is estimated in a triple way: as a total time, the time needed to download and standardise data from the phone, and the time of calculating predictions from random trees. The prediction class is the one that gets the most indications. The obtained final results are presented in 5.1.

Graph 5.2 shows two screens seen by the participant of the technical test. The tests were carried out in October and November 2021. Recruitment for technical tests took place through volunteers recruited on social networks (FB, Plazza, LinkedIn). The participant downloaded the application from the link to the resources in the cloud (via QR code) and consecutively accepted three consents, specified gender and age, and started calculations. After 1-9 seconds (the minimum and maximum of the measurements), he/she obtained a complete personality profile with the probability for each class.

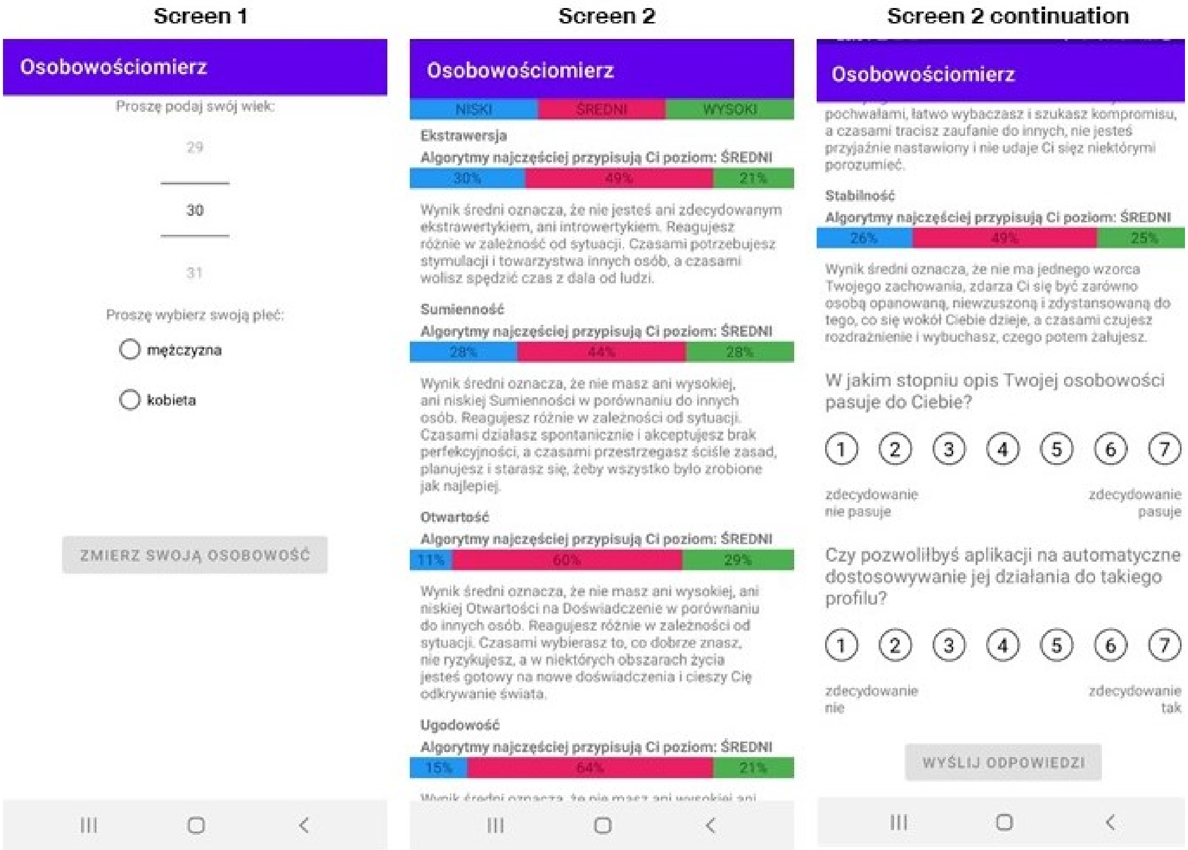


Figure 5.2. Technical App Screens (In Polish language)

(Source: own work)

In Table 5.2 pivotal values from time distribution are presented. Based on 170 phones with an Android operational system (version 8.0 or higher), the average total time needed to obtain a profile is 2.99 seconds (after excluding the weakest Samsung A6 model, the average model counting time drops below 2 seconds). Most of the time (1.8 seconds on average) takes data processing, and only about one second on average takes the execution of the prediction calculation from random trees. For evaluating the speed of micro-services, usually, two metrics are perceived as valuable: Mode and 95th percentile (Mostofi et al., 2021). In the case of the UISPP model, it can be concluded that these two measures are satisfactory. Based on the total sample in case of 95% of users, the UISPP model will be available for use in the PAPA method within 4.6 seconds; the dominant time can be close to 1.4 seconds. This result is good enough for the UISPP to use it during the app installation procedure. The speed of processors in smartphones is constantly improving, and looking at the graph 5.1, the latest generation phones

Table 5.2. Descriptive Values (in seconds) of model calculating time distribution from the test measurement on 170 smartphones and after cut off the weakest smartphone Samsung A6 from the sample. *Time total* is for global time of calculation,*time data* for time of data processing and *time model* is for calculating results from 5 traits with UISPP. (N=165)

	All smartphones N=170			Smartphones without Samsung A6, N=165		
	Time total	Time data	Time model	Time total	Time data	Time model
95th percentile	4.60	3.73	1.73	4.06	3.50	0.94
Mean	2.99	1.81	1.18	1.78	1.59	0.20
Mode	1.43	0.91	0.03	1.43	0.91	0.03
Median	1.53	1.37	0.07	1.51	1.34	0.07

will be able to count the UISPP profile in under 1 second. The test measurement does not consider non-android operating systems used in smartphones. However, the time needed to obtain the model from the data is so short that the UISPP Model can be considered efficient under the conditions required by the business (the speed of calculating the personality profile to obtain the profile during application installation).

During the technical tests, users were asked to what extent the personality profile, calculated from the data on the phone, presented with a description, matched them (5.2).

The short description referred to the most common observable behaviours for a given trait level. The chart 5.3 shows the results obtained for 129 people, and table 5.3 shows also the distribution of answers given in the groups of women and men. Most of the participants to some extent accept (answers: 5,6,7) the profile calculated by the model (67% compared to 22% ambivalent and 11% non-accepting (answers: 1, 2, 3)). According to the data, results for women and men are similar.

Additionally, the difference between people having any trait in UISPP on a high or low level versus the group with all traits on medium level is checked. In the group of people with at least one trait at a low or high level (N=36), 69% rated the profile above the middle grade on the scale. In the group of people with all features at the average level (N = 92), the analogous group accepting the profile consist of 66%. The results in the two groups did not differ significantly.

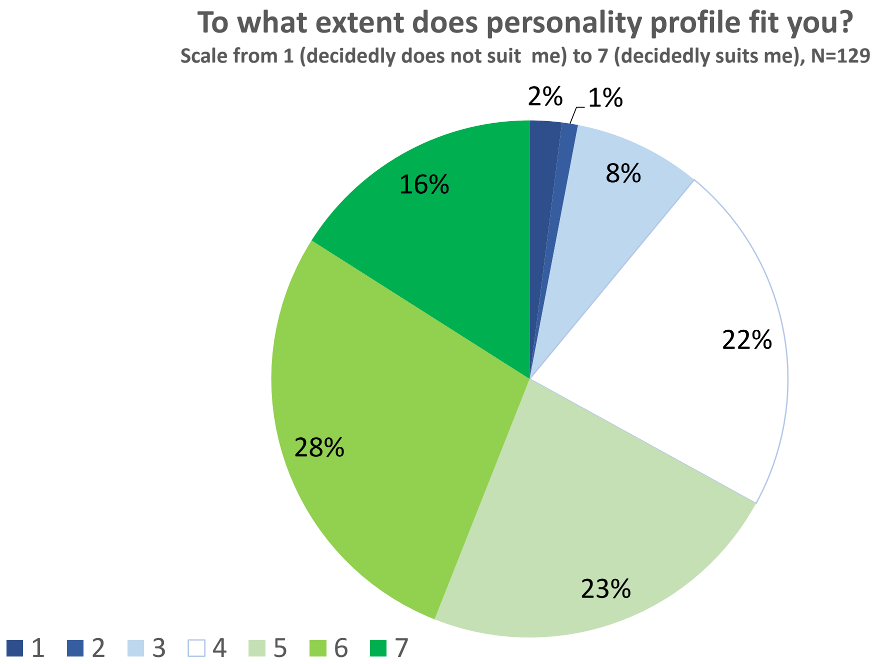


Figure 5.3. Customer Acceptance of Personality profile calculated by UISPP model. Results from user test, N=129

(Source: own work)

Table 5.3. Distribution of answers for question "To what extent does your personality profile fit you?" Answers on Likert scal: 1 - does not fit me at all; 7- 1 - decidedly fits me. Distribution among differently defined groups, with additional aggregation of negative (1-3), ambivalent (4) and positive answers (5-7) . Data from technical test. (own work)

LIKERT scale	Total		Women		Men		At least one High or Low trait		All traits on Medium level	
1	2%		0%		4%		3%		2%	
2	1%	11%	0%	12%	1%	11%	3%	14%	0%	10%
3	8%		12%		6%		8%		9%	
4	22%	22%	20%	20%	23%	23%	17%	17%	24%	24%
5	23%		27%		21%		25%		23%	
6	28%	67%	31%	68%	26%	66%	27%	69%	28%	66%
7	16%		10%		19%		17%		15%	
mean:	5.1		5.1		5.1		5.1		5.1	
N=	129		51		78		36		93	

5.3 PAPA Method Evaluation

Chapter 2.5 discusses the literature examples of the use of personality in the context of personalisation. The context of the tasks in which personality was used has always been of great importance for the results of the research. Even the very definition of personalisation proposed by Riecken (2000) (Section 2.5) talks about adapting personalisation to the context in which the user's needs are embedded. The PAPA method proposed in this dissertation cannot yet be tested in the final market service because such implementation has not yet taken place. However, the developed Android application for technical tests validates the PAPA method by criteria for meeting certain features by the method. These criteria result from the dissertation thesis itself.

In the previous chapter, the UISPP model was evaluated. Technical tests have shown that regardless of the model and computing power of the smartphone, the model counts in a short enough time to be used from the moment the user on-boards the service. It was a necessary condition for the feasibility of the PAPA method, and it has been met. Another feature of the feasibility of the PAPA method is the complete protection of user data. This condition was also met by developing the application with the UISPP model. The creation of such an application, in which all data processing and counting occur locally, is also proof of ensuring the physical protection of the user's privacy, both for his/her sensitive and personal data and any information generated on these data. It validates the method's usability despite not sharing user data and profiles derived from this data with the service provider. The UISPP model indicated people dedicated to personalisation using the PAPA method during the technical tests. The Table 5.4 shows the shares obtained during application test.

Table 5.4. Distribution of user's types in the technical app with UISPP experiment, N=170.

User's type	number	percentage
Two traits in Low class	1	1%
One trait in Low class	27	16%
All traits in Medium class	125	73%
One trait in High class	16	9%
One trait on High and one in Low class	1	1%
Total	**170**	100%

The model detected 45 people out of 170, 26.4% with distinct needs other than the typical user. In theory, this would provide an opportunity to improve the user experience via personalisation for those 26.4% of the people who installed the application. The rest of 73.4% would receive a non-personalised product. It is also worth noting how the distribution shown in the table is consistent with the distribution for personality traits in which typical users, defined by statistics as + -1SD from the average, account for about 72% of the population.

During the test also the general attitudes to personalisation based on the personality profile calculated from the data on the phone was assessed.

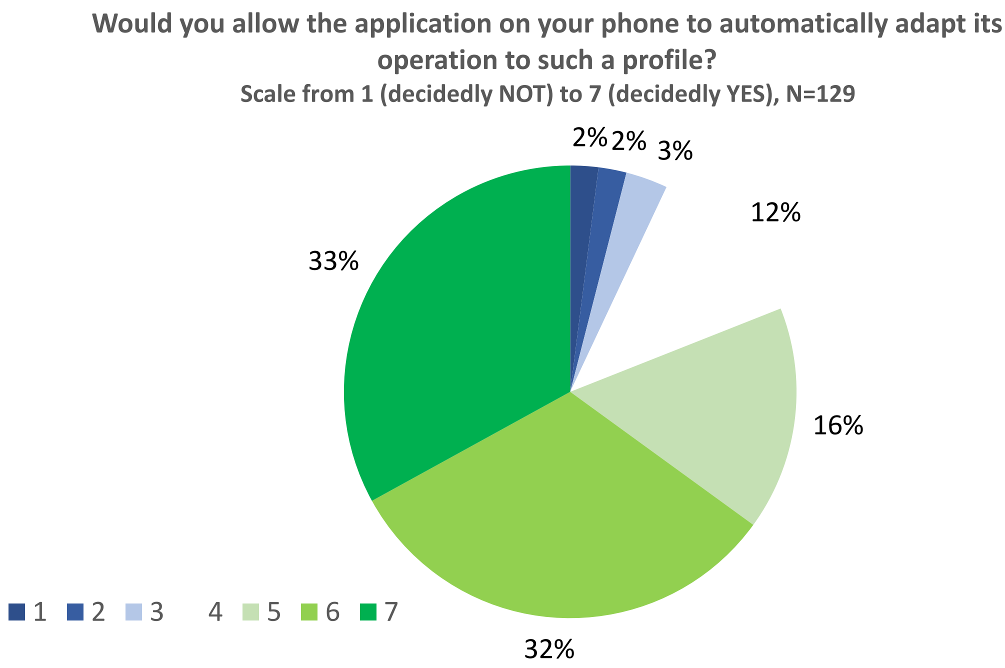


Figure 5.4. Customer Acceptance for Personalisation via PAPA method based on the profile calculated by UISPP model. Results from user test, N=129

(Source: own work)

The Figure 5.4 and the 5.5 shows the distribution of grades obtained. 81% of 129 testers who answered the question about personalisation based on the UISPP model gave scores above the ambivalent note (5-7). According to these declarations, women appear to be more likely to agree to automatic personalization based on their data-based personality profile (90% in comparison to 75% for men). According to these declarations, women appear to be more likely to agree to personalization. Unfortunately, on the basis of this question alone, it is difficult to explain why they assessed the personality profile similarly to men, and yet they are more often willing to allow for personalization. One hypothesis that may be taken into account in future research examining this paradox is, the tendency of less trust in technological skills among women. Perhaps the very term "automatic" was encouraging for them.

Table 5.5. Distribution of answers for question "Would you allow the application on your phone to automatically adapt its operation to such a profile" Answers on Likert scal: 1 - does not fit me at all; 7- 1 - decidedly fits me. Distribution among differently defined groups, with additional aggregation of negative (1-3), ambivalent (4) and positive answers (5-7). Data from technical test. (own work)

LIKERT scale	Total		Women		Men		At least one High or Low trait		All traits on Medium level	
1	2%		2%		1%		3%		1%	
2	2%	7%	2%	6%	3%	8%	0%	6%	3%	7%
3	3%		2%		4%		3%		3%	
4	12%	12%	4%	4%	17%	17%	6%	6%	14%	14%
5	16%		18%		14%		17%		15%	
6	32%	81%	39%	90%	27%	75%	43%	88%	27%	79%
7	33%		33%		34%		28%		37%	
mean:	5.8		5.8		5.6		5.8		5.7	
N=	129		51		78		36		93	

5.4 Summary of Artefacts Evaluation

The evaluation process aimed to validate the PAPA personalisation method proposed in the dissertation through the UISPP personality profile calculated from data on a smartphone. During the creation of the UISPP, the solution was constantly evaluated according to the defined criteria (chapter 5.1.2). Building the final model assumed the maximisation of parameters such as Accuracy, Precision, or a combined indicator for Precision and Recall called the F1 score (C.1 for UISPP assessment). The qualitative measures of the model reflected the degree of compliance of the predicted profile with the profile obtained in the personality survey. The obtained quality parameters were compared with variously defined benchmarks. The accuracy of this prediction is of crucial importance, and the quality of PAPA personalisation depends on the quality of this prediction. Another critical element is adjusting the speed at which the profile is computed. In order to be able to implement the personalisation at the earliest possible stage of the service, the time needed to calculate it can not be longer than a few seconds (C.2 for UISPP assessment). During the technical test, users were asked to evaluate the personality profile's accuracy based on the data (C.3 for the UISPP assessment). While the first two criteria (C.1 and C.2) were crucial for using the UISPP profile in the PAPA method, criterion C.3 was treated only as supplementary information from users, proving their Openness to the proposed solution and did seem too critical. The PAPA evaluation was made based on the feasibility of the UISPP (C.4) method as it works automatically in a real-world environment (android mobile application) used for user tests. The results of these tests also confirmed the ability of the method to work, despite the use of local data processing only, which does not require sharing user data and profiles with the service provider (C.5). Moreover, as in the case of UISPP, this solution was assessed by the users-testers (C.6). The results of the UISPP and PAPA evaluation can be summarised in the following points:

1. In the development and evaluation cycle, the UISPP model with the best quality parameters was selected. The aim was to obtain a model not worse than other predictive models but meeting the technical criteria for processing local data (privacy) and processing only data found on the phone at the time of the installation of the service. The conducted experiments proved that despite the limitations applied in the construction of the model (data available on the phone) and limitations in the shape of the model itself (number of trees), the parameters for the quality of prediction are not worse than those obtained from other data (without the limitations adopted for UISPP). The models for Extraversion, Stability and Agreeableness achieved the best results in the evaluation. The models for Conscientiousness and Openness obtained worse, though acceptable results.
2. In the end-user tests, the implementation of the UISPP in a smartphone application was first shown to be feasible. The results obtained from many popular smartphone models give grounds to state that regardless of the model, it is possible to calculate the personality profile with UISPP in a sufficiently short time during the installation of the application. A user personality profile applicable for personalisation (PAPA model) can be obtained without sharing personal and sensitive information. The profile can also be protected and available only for a given application. Moreover, the profile can be calculated based on a relatively small

amount of data without the necessity of monitoring the behaviour in time.

3. Based on individual questions, testers' opinions were also verified regarding the presented personality profile and the concept of using this automatically calculated profile for application personalisation. It can be concluded that the vast majority of testers have more or less accepted or assessed neutrally both the profile included by UISPP and the concept of personalising the application based on this profile. It should be noted that such a measurement is entirely artificial and does not take place in the situation of the user's natural contact with the application, where the user will not be asked to accept the profile but rather to consent required by law to use the data on the phone for personalisation. This consent will be necessary before the app counts the UISPP.
4. During technical tests on user-testers, it was confirmed that the UISPP model could detect people with special needs, for whom the personalisation of services may be much more critical than for about 70% of typical users. The evaluation measurement identified 45 people with special needs out of 170 test participants.

To sum up, based on verification on a sample application installed on smartphones, the performed evaluation supports this dissertation's thesis, that automatic identification of the user's personality profile based on the data available during the installation of the electronic service is possible and practical. Furthermore, this identification can also be used as a classifier facilitating the personalisation of this service.

Chapter 6

Summary

From the perspective of the literature review conducted for this dissertation, the research and its results contributed to enriching the existing knowledge base in several aspects and domains of social sciences.

First, the possibility of data-based assigning a smartphone user to classes of five personality traits (Big 5 theory) was examined. In the conducted literature review, no similar project was found (which does not exclude the existence of similar solutions in the business area). The presented work defines research gap in determining personality profiles from data found on a smartphone and in automated data-based solutions enabling the individualisation of user experience almost from the first contact with any service available through a smartphone application.

What is essential and constitutes the primary goal of the research, it is that such a profile could be calculated based on ML techniques in a short time, based on data available on the phone without the need to collect it (which is the basis of the existing solutions). Moreover, this calculation is done despite the data scope being limited to the easily available data on the smartphone, without tracking users' behaviour. Despite this limitation, the preparation of a sufficiently reliable method of quick classification of users who may need a different experience with the service is accomplished.

The literature review shows that personalisation is currently an essential strategy for building a competitive advantage in lean or agile enterprises. Quick access to the user's profile and the possibility of automating the individualisation of the customer's experience with the service seems to be perfectly suited to this business strategy. In recognition of the conducted research, it was indirectly shown that extreme personality traits differentiate the way users use a smartphone, contributing to behavioural psychology and individual differences psychology. In addition, the demonstration of the UISPP model and its implementation in PAPA showed that thanks to the limitation of end device data processing (Edge Computing), it is possible to protect user data fully.

The main advantage of PAPA is the confidentiality of the data necessary to create the user's personality profile (as it remains in the private sphere) and the profile itself. The solution facilitates better protection of the user's privacy than the existing solutions, where data processing, profile counting and storage of the profile itself takes place outside the end device on external resources (servers, cloud solutions). The obtained results allow for the conclusion that it is possible to achieve a solution not worse than others, despite the application of several limitations and additional conditions related to the protection and local processing of data. The obtained models, which classify users according to the Big 5 personality profile, are sufficiently reliable and can be used as the basis for personalising any service available through a smartphone application. In addition, the work carried out contributed to the analysis of the methods of creating analogous models, showing the differences that can be achieved in non-binary classification for five personality traits using different algorithms. It should be emphasised that the proposed models should be developed and improved. The emergence of new techniques for implementing ML models directly on the phone or improving the smartphone's capabilities is a reason to obtain better results. The continuously increasing the computing power of the smartphones themselves may soon allow the inclusion of statistics from photos in the creation of models. According to existing research, the photo data implementation in initial dataset, could improve the model parameters , especially for features such as Openness for Experience,

The method proposed in the dissertation deals with the possibility of automatic personalisation of the service (on the example of a smartphone application), based on the user's personality profile calculated from the available data on the smartphone. The research focused on creating and implementing the UISPP model in an Android application. In the context of the data used, the environment and assumptions, the results are innovative and seem to be possible to implement in any service provided to the client in the form of a smartphone application. The primary goal for creating UISPP and PAPA was implementing the user-centric services approach, which is key to creating a competitive advantage in the era of lean and agile enterprises.

First of all, the research done for the dissertation should be continued and tested on higher samples of users. The critical aspect also seems to be the confirmation of the model's universality over time or proposing modifications that allow for the UISPP validity verification (one of the assumptions for PAPA).

Development in the area of the proposed solution. The very adaptation of the service (interface, communication, organising and presenting elements or creating different affordances) requires extensive separate research checking both the preferences and their universality between different services. The aim of such research should be to gather knowledge on how to adapt services to different user's profiles and test hypotheses regarding the importance of the impact of individual personality traits of the user on the customer experience.

It would be challenging to design a self-adapting service using PAPA without additional research in these areas. Checking the primary hypotheses about dependencies and the factors that modify them is necessary to create automation of service personalisation.

An interesting aspect that requires expanding is the change of these preferences over time, such as getting used to the service, learning it, and being an experienced user. An additional critical aspect related to the PAPA method is continuing research in creating an automatic system enabling data-based evaluation of user satisfaction or dissatisfaction with the service. The primary and obvious criterion is customer loyalty(no churn). However, it seems that it would be possible to enrich the service with early warning systems for resentment or dissatisfaction and to create a system that would allow the assessment of acceptance, and satisfaction (or dissatisfaction) with individual service functionalities, e.g. using behavioural biometrics and the analysis of differences in touch screen operation.

In addition, the literature review shows that models based on data from behaviour registers and other types of data, e.g. photos, can significantly improve the accuracy of calculating the personality profile. Therefore, an exciting direction would be researching the algorithms proposed in the PAPA project that check the compliance of UISPP with information about the user obtained during the use of the service.

Finally, technologies are rapidly evolving, and smartphones are enriched with new functions and wearable devices (smartwatches) or devices in an intelligent home. Research aimed at exploring the possibilities of improving the validity of the UISPP model based on new data available on the phone is, therefore, a natural direction. Perhaps faster processors and parameters of the smartphones themselves will soon include data from photos rejected in the dissertation due to en overly long a processing time (contrary to the adopted conditions for the UISPP model).

Another application of the UISPP and PAPA model. The introduced limitations when creating the solution (complete data protection and local processing on the end device) allow for the extension of the use of the model itself for purposes other than personalisation of user experience with the service, for example, in the work environment, as a handy tool for classifying employees in terms of their essential needs and natural motivators.

Recommendation systems are another area that seems to benefit from automatic user personality classifiers. Based on the personality theory and the description of preferences for low and high levels of traits according to the Big Five, it seems that profile knowledge can significantly better adjust content recommendations, e.g. on film or television platforms. Another exciting area that can benefit from such a model are systems that analyse users' emotions. Emotionality is undoubtedly correlated with temperament, and therefore with the Big Five profile. The correct detection of the user's emotions, becoming more and more common, can significantly improve the quality of the analysis, and has a crucial impact on systems that respond to emotions and emotional changes. By definition, neurotic people are much more often emotionally unstable, which significantly influences the assessment of their emotional state (differentiation of individual baseline patterns).

In the personality literature review, much research is conducted on matching the personality of virtual entities (bots, agents, virtual assistants) and humanoid robots to the user's personality. It seems that for this purpose, having a user/caller profile from the first contact is an essential requirement. The traditional questionnaires measure personality in laboratory tests or research conducted in the work environment. However, in the case of the increase in the popularity of robots and their use for professional roles, e.g. receptionists, secretaries, assistants, in which contacts are one-off and random. The automatic methods of determining the personality profile will be of growing importance. Automatic methods of calculating the personality profile from data may also gain importance in e-health and e-work. The human personality is crucial for defining motivational goals as well, as it allows to predict problems that may arise during long-term convalescence or long-term processes. The simplest example is that a low conscientious person needs to be better monitored when taking medications or organising work. In order to function better, an introvert should have limited and monitored stress related to contact with people. Emotionally unstable people are particularly vulnerable to low mood, discouragement, depression and even suicide when stressed too much. Adapting the method of rehabilitation or occupational therapy to the personality may affect its effectiveness. For example, the therapy exercises should be characterized by high stability and repeatability for highly conscientious and low open people. In turn, it should be variable and not monotonous for people with high openness who do not tolerate repeatability and the lack of new stimuli.

Limitations. The proposed method currently has two fundamental limitations. The first one is related to the fact that the prototype was built in the Android system, and due to overly high implementation costs, it has not been verified whether it is fully reproducible for other systems, e.g. for IOS (Apple). It probably will not be done until the UISPP and PAPA implementation in the final market service. This dissertation was not checked either whether there is the same possibility of using transpilers of Random Forest trees to the iOS system. According to data from the Global Stats website, "in May 2020, 99.3% of smartphones on the market are equipped with one of two operating systems - Android, whose market share is over 72.5%, and iOS, which supports 26.8% of devices." [1].

Possible implementation in the commercial service will require additional research and possible adaptation with modifications of the solution from Android to iOS. The second fundamental limitation is a particular case where the user installs the application - a service on an entirely new smartphone that has not yet been used or cloned settings and data from the previously used phone. For example, this problem affects children and teenagers who get their first smartphone. Unfortunately, it is impossible to count the profile in such a situation. In such a situation, UISPP calculation, and thus the PAPA method, will be possible only after some time has elapsed.

[1] cite from https://www.orange.pl/poradnik/smartfony-i-inne-urzadzenia/czy-istnieje-zycie-poza-ios-i-androidem-telefon-z-wlasnym-systemem-operacyjny/

Bibliography

Alam, F., Stepanov, E. A., & Riccardi, G. (2013). Personality traits recognition on social network-facebook. *Proceedings of the International AAAI Conference on Web and Social Media*, *7*(1).

Allik, J., & McCrae, R. R. (2006). Erratum: Toward a geography of personality traits: Patterns of profiles across 36 cultures (Journal of Cross-Cultural Psychology). https://doi.org/10.1177/0022022106288482

Allik, J., & Mccrae, R. R. (2004). Toward a Geography of Personality Traits: Patterns of Profiles across 36 Cultures. *Journal of Cross-Cultural Psychology*. https://doi.org/10.1177/0022022103260382

Allport, G. W. (1937). Classics in the History of Psychology – Allport (1937). *Classics in the History of Psychology –Allport*.

Amnesty_International. (2019). Surveillance giants: How the business model of Google and Facebook threatens human rights.

Anand, P. (2009 (accessed August 30, 2019)). The Reality Behind Voice Shopping Hype.

Angleitner, A. (1991). Personality psychology: Trends and developments. *European Journal of Personality*, *5*, 185–197.

Armstrong, M. (2007). A Handbook of Management Techniques: A Comprehensive Guide to Achieving Managerial Excellence and Improved Decision Making. *Leadership & Organization Development Journal*, *28*(4).

Arnoux, P.-H., Xu, A., Boyette, N., Mahmud, J., Akkiraju, R., & Sinha, V. (2017). 25 tweets to know you: A new model to predict personality with social media. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1).

Arroju, M., Hassan, A., & Farnadi, G. (2015). Age, gender and personality recognition using tweets in a multilingual setting. *6th Conference and Labs of the Evaluation*

Forum (CLEF 2015): Experimental IR meets multilinguality, multimodality, and interaction, 23–31.

Asensio, J. A., Criado, J., Padilla, N., & Iribarne, L. (2019). Emulating home automation installations through component-based web technology. *Future Generation Computer Systems*, *93*. https://doi.org/10.1016/j.future.2017.09.062

Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS quarterly*, 13–28.

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. https://doi.org/10.1016/j.paid.2017.12.018

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*. https://doi.org/10.1177/0956797609360756

Ball, D., Coelho, P. S., & Vilares, M. J. (2006). Service personalization and loyalty. *Journal of services marketing*.

Barrick, M. R., & Mount, M. K. (1991). The Big Five Personality Dimensions And Job Performance: A Meta-Analysis. *Personnel Psychology*. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Becker, B. W., & Connor, P. E. (1981). Personal values of the heavy user of mass media. *Journal of Advertising research*.

Bedingfield, C. (2003). A pattern language for web usability. *Ubiquity*, *2003*(August). https://doi.org/10.1145/964680.941400

Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J., & Kleitman, S. (2019). Detecting personality traits using eye-tracking data. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300451

Bharadwaj, S., Sridhar, S., Choudhary, R., & Srinath, R. (2018). persona traits identification based on myers-briggs type indicator (MBTI)-a text classification approach. *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1076–1082.

Bianchi, A., & Phillips, J. G. (2005). Psychological predictors of problem mobile phone use. *Cyberpsychology and Behavior*. https://doi.org/10.1089/cpb.2005.8.39

Bin Tareaf, R., Alhosseini, S. A., & Meinel, C. (2019). Facial based personality prediction models for estimating individuals private traits. *Proceedings - 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA,BDCloud,SustainCom,SocialCom 2019*. https://doi.org/10.1109ISPA-BDCloud-SustainCom-SocialCom48970.2019.00233

Blom, J. (2000). Personalization: A Taxonomy, 313–314. https://doi.org/10.1145/633292.633483

Boyle, G. J., Matthews, G., & Saklofske, D. H. (2008). *The SAGE handbook of personality theory and assessment: Volume 1 — personality theories and models*. https://doi.org/10.4135/9781849200462

Božič, K., & Dimovski, V. (2019). Business intelligence and analytics for value creation: The role of absorptive capacity. *International journal of information management*, *46*, 93–103.

Branco, S., Ferreira, A. G., & Cabral, J. (2019). Machine Learning in Resource-Scarce Embedded Systems, FPGAs, and End-Devices: A Survey. *Electronics*, *8*(11). https://doi.org/10.3390/electronics8111289

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Buraya, K., Farseev, A., Filchenkov, A., & Chua, T.-S. (2017). Towards user personality profiling from multiple social networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1).

Carlo, G., Knight, G. P., Roesch, S. C., Opal, D., & Davis, A. (2013). Personality across cultures: A critical analysis of Big Five research and current directions. *APA handbook of multicultural psychology, Vol. 1: Theory and research*. https://doi.org/10.1037/14189-015

Celli, F. (2011). Mining user personality in twitter. *Language, Interaction and Computation CLIC*.

Celli, F. (2012). Unsupervised personality recognition for social network sites. *Proc. of sixth international conference on digital society*, 59–62.

Celli, F., & Rossi, L. (2012). The role of emotional stability in Twitter conversations. *Proceedings of the workshop on semantic analysis in social media*, 10–17.

Chaudhary, S., Singh, R., Hasan, S. T., & Kaur, M. I. (2013). A comparative Study of Different Classifiers for Myers-Brigg Personality Prediction Model. *Linguistic analysis*, 21.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Cheung, F. M., Leung, K., Zhang, J.-X., Sun, H.-F., Gan, Y.-Q., Song, W.-Z., & Xie, D. (2001). Indigenous Chinese Personality Constructs. *Journal of Cross-Cultural Psychology*. https://doi.org/10.1177/0022022101032004003

Cheverst, K., Mitchell, K., & Davies, N. (2002). The role of adaptive hypermedIa in a context-aware tourist guide. *Communications of the ACM*, *45*(5). https://doi.org/10.1145/506218.506244

Chishti, S., Li, X., & Sarrafzadeh, A. (2015). Identify Website Personality by Using Unsupervised Learning Based on Quantitative Website Elements. *International Conference on Neural Information Processing*, 522–530.

Clarke III, I. (2008). Emerging value propositions for m-commerce. *Journal of Business Strategies*, *25*(2).

Cody, K. (2012). “BeTween two worlds”: critically exploring marketing segmentation and liminal consumers. *Young Consumers*.

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.38.4.668

Costa, P. T., & McCrae, R. R. (1992). Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory. *Psychological Assessment*. https://doi.org/10.1037/1040-3590.4.1.5

Costa, P. T., & McCrae, R. R. (1995). Domains and Facets: Hierarchical Personality Assessment Using the Revised NEO Personality Inventory. *Journal of Personality Assessment*. https://doi.org/10.1207/s15327752jpa6401-2

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management science*, *35*(8), 982–1003.

De Raad, B. (1999). Interpersonal Lexicon: Structural Evidence from Two Independently Constructed Verb-Based Taxonomies. *European Journal of Psychological Assessment*. https://doi.org/10.1027//1015-5759.15.3.181

De Raad, B. (2012). Structural models of personality. *The Cambridge Handbook of Personality Psychology*. https://doi.org/10.1017/cbo9780511596544.011

DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*. https://doi.org/10.1016/j.jrp.2014.07.004

Digman, J. M. (1990). Personality structure: emergence of the five-factor model. *Annual Review of Psychology*. https://doi.org/10.1146/annurev.ps.41.020190.002221

Drucker, P. (2012). *The Practice of Management*. https://doi.org/10.4324/9780080942360

Egger, R., & Rainoldi, M. (2021). Co-creating Personalised Experiences in the Context of the Personalisation-Privacy Paradox. *Information and Communication Technologies in Tourism 2021: Proceedings of the ENTER 2021 ETourism Conference, January 19-22, 2021*, 95.

Eysenck, H. J. (1981). General Features of the Model. *A Model for Personality*. https://doi.org/10.1007/978-3-642-67783-0-1

Eysenck, H. J. (2019). Trait Theories Of Personality. *Companion Encyclopedia of Psychology*. https://doi.org/10.4324/9781315002897-36

Fan, H., & Poole, M. (2006). What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems. *Journal of Organizational Computing and Electronic Commerce - J ORGAN COMPUT ELECTRON COMME*, *16*, 179–202. https://doi.org/10.1207/s15327744joce1603&4-2

Feliksiak, M. (2016). Inwigilacja w internecie.

Finck, M. (2021). The Limits of the GDPR in the Personalisation Context.

Fiorentini, N., & Losa, M. (2020). Handling imbalanced data in road crash severity prediction by machine learning algorithms. *Infrastructures*, *5*(7), 61.

Frey, R. M., Xu, R., Ammendola, C., Moling, O., Giglio, G., & Ilic, A. (2017). Mobile recommendations based on interest prediction from consumer's installed apps–insights from a large-scale field study. *Information Systems*, *71*, 152–163. https://doi.org/10.1016/j.is.2017.08.006

Gjurković, M., Karan, M., Vukojević, I., Bošnjak, M., & Šnajder, J. (2020). PANDORA talks: Personality and demographics on reddit. https://doi.org/10.31234/osf.io/94xcp

Gjurković, M., & Šnajder, J. (2018). Reddit: A gold mine for personality prediction. *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, 87–97.

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from twitter. *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, 149–156. https://doi.org/10.1109/PASSAT/SocialCom.2011.33

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*.

Goldberg, L. R. (1992). The Development of Markers for the Big-Five Factor Structure. *Psychological Assessment*. https://doi.org/10.1037/1040-3590.4.1.26

Hagen, P., Manning, H., & Souza, R. (1999). Smart personalization. *Forrester Research, Cambridge, MA*.

Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*. https://doi.org/10.1207/s15327906mbr4003_2

Hall C.S., C. J. B., Lindzey G. (2004). *Theories of Personality, 4th Edition*.

Hevner, A., & Chateerjee, S. (2012). Integrated Series in Information Systems Volume 28. https://doi.org/10.1007/978-1-4419-6108-2

Hevner, A., & Chatterjee, S. (2010). Design Science Research in Information Systems. https://doi.org/10.1007/978-1-4419-5653-8-2

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly: Management Information Systems*. https://doi.org/10.2307/25148625

Hinds, J., Williams, E. J., & Joinson, A. N. (2020). "It wouldn't happen to me": Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*, *143*, 102498.

Hoek, P., Peelen, P., & Commandeur, P. (1999). Achieving Mass Customization Through Postponement: A Study of International Changes. *Journal of Market-Focused Management*, *3*(3-4). https://doi.org/10.1023/A:1009827812231

Hofstee, W. K. B. (1991). The Concepts of Personality and Temperament. In J. Strelau and A. Angleitner (Eds.), *Explorations in Temperament: International Perspectives on Theory and Measurement* (pp. 177–188). Springer US. https://doi.org/10.1007/978-1-4899-0643-4_12

Hornowska, E. (2018). *Testy psychologiczne: teoria i praktyka*. Wydawnictwo Naukowe" Scholar".

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1–55.

Huda, S., Liu, K., Abdelrazek, M., Ibrahim, A., Alyahya, S., Al-Dossari, H., & Ahmad, S. (2018). An ensemble oversampling model for class imbalance problem in software defect prediction. *IEEE access*, *6*, 24184–24195.

Idberg, L., Orfanidou, S., & Karppinen, O. (2021). Privacy for sale!: An exploratory study of personalization privacy paradox in consumers' response to personalized advertisements on social networking sites.

Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, *51*(8), 56–59.

Isabel Myers, P. M. (2010). *Gifts Differing: Understanding Personality Type*.

Ives, B., & Piccoli, G. (2003). Custom Made Apparel and Individualized Service at Lands' End. *Communications of the Association for Information Systems*, *11*(1), 3.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five Trait taxonomy. https://doi.org/10.1016/S0191-8869(97)81000-8

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*. https://doi.org/10.1111/j.1744-6570.1999.tb00174.x

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*. https://doi.org/10.1037/0021-9010.87.4.797

Kalimeri, K., Lepri, B., & Pianesi, F. (2013). Going beyond traits: Multimodal classification of personality states in the wild. *ICMI 2013 - Proceedings of the 2013 ACM International Conference on Multimodal Interaction*, 27–34. https://doi.org/10.1145/2522848.2522878

Kalyanaraman, S., & Sundar, S. (2006). The psychological appeal of personalized content in web portals: Does customization affect attitudes and behavior? *Journal of Communication*, *56*(1). https://doi.org/10.1111/j.1460-2466.2006.00006.x

Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., Kosinski, M., Ramones, S. M., & Seligman, M. E. (2014). The Online Social Self: An Open Vocabulary Approach to Personality. *Assessment*. https://doi.org/10.1177/1073191113514104

Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Park, G., Ungar, L. H., Stillwell, D. J., Kosinski, M., Dziurzynski, L., & Seligman, M. E. (2014). From "sooo excited!!! "to "so proud": Using language to study development. *Developmental Psychology*. https://doi.org/10.1037/a0035048

Khan, A. S., Ahmad, H., Asghar, M. Z., Saddozai, F. K., Arif, A., & Khalid, H. A. (2020). Personality classification from online text using machine learning approach. *International Journal of Advanced Computer Science and Applications*, *11*(3). https://doi.org/10.14569/ijacsa.2020.0110358

Kidd, P. (1994). *Agile Manufacturing: Forging New Frontiers*. Addison-Wesley.

Kosinski, M., Bachrach, Y., Kasneci, G., Van-Gael, J., & Graepel, T. (2012). Crowd IQ: Measuring the intelligence of crowdsourcing platforms. *Proceedings of the 4th Annual ACM Web Science Conference, WebSci'12*. https://doi.org/10.1145/2380718.2380739

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kotler, P., & Turner, R. E. (1997). *Marketing management: Analysis, planning, implementation, and control* (Vol. 9). Prentice hall Upper Saddle River, NJ.

Kowalski, A. M., & Weresa, M. A. (2019). RAPORT O KONKURENCYJNOŚCI 2019. *RAPORT O KONKURENCYJNOŚCI 2019*, 247.

Kreuter, M. W., & Wray, R. J. (2003). Tailored and targeted health communication: strategies for enhancing information relevance. *American journal of health behavior*, *27*(1), S227–S232.

Krzeminska, I. (2020). Personalizing smart services based on data-driven personality of user. *11878 LNCS*.

Krzeminska, I. (2019). Personalizing Smart Services Based on Data-Driven Personality of User. *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, 199–203.

Krzeminska, I., & Rzeznik, J. (2021). Personality-Based Lexical Differences in Services Adaptation Process. *Technium: Romanian Journal of Applied Sciences and Technology*, *3*(1), 61–73.

Krzemińska, I., & Szmydt, M. (2021). Personality Based Data-driven Personalization as an Integral Part of the Mobile Application. *LNCS*.

Kumar, S., & Desai, D. (2016). Web personalization: a perspective of design and implementation strategies in websites. *KHOJ: Journal of Indian Management Research and Practices*, 109–119.

Kurzweil, R. (2004). The Law of Accelerating Returns. *Alan Turing: Life and Legacy of a Great Thinker*. https://doi.org/10.1007/978-3-662-05642-4_16

Larsen, S., & Tutterow, S. (1999). Developing the Personalization-Centric Enterprise Through Collaborative Filtering and Rules-based Technologies. *CRM Project*, *1*.

Laterza, V. (2018). Cambridge Analytica, independent research and the national interest. *Anthropology Today*, *34*(3), 1–2.

Levy, Y., & J. Ellis, T. (2006). Towards a Framework of Literature Review Process in Support of Information Systems Research. *Proceedings of the 2006 InSITE Conference*. https://doi.org/10.28945/2980

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2017). Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, *18*(1), 6765–6816.

Li, Y., Kazameini, A., Mehta, Y., & Cambria, E. (2021). Multitask Learning for Emotion and Personality Detection.

Liu, L., Preoţiuc-Pietro, D., Samani, Z. R., Moghaddam, M. E., & Ungar, L. (2016). Analyzing personality through social media profile picture choice. *Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016*, 211–220.

Lukito, L. C., Erwin, A., Purnama, J., & Danoekoesoemo, W. (2016). Social media user personality classification using computational linguistic. *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 1–6.

Maes, P., Guttman, R. H., & Moukas, A. G. (1999). Agents that buy and sell. *Communications of the ACM*, *42*(3). https://doi.org/10.1145/295685.295716

Magnusson, D. (1967). *Test Theory: Transl. from the Swedish by Hunter Mabon*. Addison-Wesley.

Magnusson, D. (1991). *Wprowadzenie do teorii testów*. Państwowe Wydawn. Naukowe.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.88.1.139

Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological methods*, *9*(3), 275.

Matthews, G. (2008). Personality and information processing: A cognitive-adaptive theory. *The SAGE Handbook of Personality Theory and Assessment: Volume 1 - Personality Theories and Models*. https://doi.org/10.4135/9781849200462.n3

Matthews, G. (2020). Against consensus: Embracing the disunity of personality theory. *Personality and Individual Differences*. https://doi.org/10.1016/j.paid.2019.109535

Matthews, G., Hancock, P. A., Lin, J., Panganiban, A. R., Reinerman-Jones, L. E., Szalma, J. L., & Wohleber, R. W. (2020). Evolution and revolution: Personality research for the coming world of robots, artificial intelligence, and autonomous systems. *Personality and Individual Differences*. https://doi.org/10.1016/j.paid.2020.109969

Matthews, G., Lin, J., Panganiban, A. R., & Long, M. D. (2020). Individual Differences in Trust in Autonomous Robots: Implications for Transparency. *IEEE Transactions on Human-Machine Systems*. https://doi.org/10.1109/THMS.2019.2947592

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

McCann, S. J. (2005). Longevity, big five personality factors, and health behaviors: Presidents from washington to nixon. *Journal of Psychology: Interdisciplinary and Applied*. https://doi.org/10.3200/JRLP.139.3.273-288

McCrae, R. R. (2002). Cross-Cultural Research on the Five-Factor Model of Personality. *Online Readings in Psychology and Culture*. https://doi.org/10.9707/2307-0919.1038

McCrae, R. R., & Costa, P. T. (1987). Validation of the Five-Factor Model of Personality Across Instruments and Observers. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.52.1.81

McCrae, R. R., & Costa, P. T. (1996). Toward a New Generation of Personality Theories: Theoretical Contexts for the Five-Factor Model. *The Five-Factor Model of Personality: Theoretical Perspectives*.

McCrae, R. R., & Costa, P. T. (1997). Personality Trait Structure as a Human Universal. *American Psychologist*. https://doi.org/10.1037%/0003-066X.52.5.509

McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*. https://doi.org/10.1016/S0191-8869(03)00118-1

McCrae, R. R., & Costa, P. T. (2012). Introduction to the empirical and theoretical status of the five-factor model of personality traits. *Personality disorders and the five-factor model of personality (3rd ed.)*. https://doi.org/10.1037/13939-002

McCrae, R. R., & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Mettler, T., Eurich, M., & Winter, R. (2014). On the use of experiments in design science research: A proposition of an evaluation framework. *Communications of the Association for Information Systems*. https://doi.org/10.17705/1cais.03410

Montjoye, Y.-a. D., Quoidbach, J., & Robic, F. (2013). Predicting personality using novel phone-based metrics. *Social Computing, Behavioral-Cultural Modeling and Prediction Lecture Notes in Computer Science*, 48–55.

Mostofi, V. M., Krishnamurthy, D., & Arlitt, M. (2021). Fast and Efficient Performance Tuning of Microservices. *2021 IEEE 14th International Conference on Cloud Computing (CLOUD)*, 515–520. https://doi.org/10.1109/CLOUD53861.2021.00067

Murray, H. A., & McAdams, D. P. (2010). *Explorations in Personality*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195305067.001.0001

Nadeski, M. (2019). Bringing machine learning to embedded systems. *Texas Instruments*.

Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E. B., & Turaga, D. S. (2017). Learning Feature Engineering for Classification. *Ijcai*, 2529–2535.

Nasiri, M., Rantala, T., Saunila, M., Ukko, J., & Rantanen, H. (2018). Transition towards sustainable solutions: Product, service, technology, and business model. *Sustainability (Switzerland)*, *10*(2). https://doi.org/10.3390/su10020358

Ngatirin, N. R., Zainol, Z., & Yoong, T. L. C. (2016). A comparative study of different classifiers for automatic personality prediction. *2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 435–440.

Nie, D., Guan, Z., Hao, B., Bai, S., & Zhu, T. (2014). Predicting personality on social media with semi-supervised learning. *Proceedings - 2014 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT 2014*, *2*. https://doi.org/10.1109/WI-IAT.2014.93

Ning, H., Dhelim, S., & Aung, N. (2019). PersoNet: Friend Recommendation System Based on Big-Five Personality Traits and Hybrid Filtering. *IEEE Transactions on Computational Social Systems*, *6*(3). https://doi.org/10.1109/TCSS.2019.2903857

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*. https://doi.org/10.1037/h0040291

N.S., H., M., S., & M.D., G. (2019). The Digital Addiction Scale for Children: Development and Validation. *Cyberpsychology, behavior and social networking*.

Okoli, C. (2015). A Guide to Conducting Systematic Literature Review. *Communications of the Association for Information Systems*.

Ones, D. S., Mount, M. K., Barrick, M. R., & Hunter, J. E. (1994). Personality and Job Performance: A Crittique Tett, Jackson and Rothstein (1991) Meta-Analysis. *Personnel Psychology*. https://doi.org/10.1111/j.1744-6570.1994.tb02414.x

Ong, V., Rahmanto, A. D., Suhartono, D., Nugroho, A. E., Andangsari, E. W., Suprayogi, M. N., et al. (2017). Personality prediction based on Twitter information in Bahasa Indonesia. *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 367–372.

Ong, V., Rahmanto, A. D., Williem, & Suhartono, D. (2017). Exploring personality prediction from text on social media: a literature review. *Internetworking Indonesia*, *9*(1), 65–70.

Pal, K., & Patel, B. V. (2020). Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques. *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00016

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., Seligman, M. E. P., Park, G., Schwartz, H. A., Kern, J. C., Kosinski, M. L., & Stillwell, M. (2014). Journal of Personality and Social Psychology Automatic Personality Assessment Through Social Media Language Automatic Personality Assessment Through Social Media Language. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000020

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*.

Peppers, D., & Rogers, M. (1993). *The one to one future: Building relationships one customer at a time*. Currency Doubleday New York.

Plank, B., & Hovy, D. (2015). Personality traits on twitter—or—how to get 1,500 personality tests in a week. *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 92–98.

Polzehl, T. (2015). *Personality in speech: Assessment and automatic classification.*

Pratama, B. Y., & Sarno, R. (2015). Personality classification based on Twitter text using Naive Bayes, KNN and SVM. *2015 International Conference on Data and Software Engineering (ICoDSE)*, 170–174.

Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011). Our twitter profiles, our selves: Predicting personality with twitter. *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*. https://doi.org/10.1109/PASSAT/SocialCom.2011.26

Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., & Crowcroft, J. (2012). The personality of popular facebook users. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. https://doi.org/10.1145/2145204.2145346

Quinn, E. (2021). Surveillance Giants: The Possibilities & Limits of The Business and Human Rights Framework in the Data Economy.

Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*, *3*(24). https://doi.org/10.21105/joss.00638

Raschka, S., Patterson, J., & Nolet, C. (2020). Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. https://doi.org/10.3390/info11040193

(red), H. B. (2012). *Logistyka przyszłości*. Polskie Wydawnictwo Ekonomiczne, Warszawa.

Riecken, D. (2000). Personalized Views of Personalization. *Communications of the ACM*. https://doi.org/10.1145/345124.345133

Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(4), 369–379.

Robert, L. P. (2018). Personality in the human robot interaction literature: A review and brief critique. *Americas Conference on Information Systems 2018: Digital Disruption, AMCIS 2018*.

Rubenzer, S. J., Faschingbauer, T. R., & Ones, D. S. (2000). Assessing the U.S. presidents using the revised NEO Personality Inventory. *Assessment*. https://doi.org/10.1177/107319110000700408

Ruch, W., Angleitner, A., & Strelau, J. (1991). The Strelau Temperament Inventory—Revised (STI-R): Validity studies. *European Journal of Personality*. https://doi.org/10.1002/per.2410050403

Sadeh, N. (2003). *M-commerce: technologies, services, and business models*. John Wiley & Sons.

Sahu, Y., Ramani, Y., Parekh, V., & Maru, N. (2019). Personality prediction from social media images: A content driven approach. *Proceedings of the 2019 6th International Conference on Computing for Sustainable Global Development, INDIACom 2019*.

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would You Trust a (Faulty) Robot? https://doi.org/10.1145/2696454.2696497

Samuel, A. L. (1959). Some Studies in Machine Learning. *IBM Journal of Research and Development*, *3*(3).

Saucier, G. (2009). Recurrent personality dimensions in inclusive lexical studies: Indications for a big six structure. *Journal of Personality*. https://doi.org/10.1111/j.1467-6494.2009.00593.x

Saulsman, L. M., & Page, A. C. (2004). The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review*. https://doi.org/10.1016/j.cpr.2002.09.001

Schäfer, I., Hansen, H., Schön, G., Höfels, S., Altiner, A., Dahlhaus, A., Gensichen, J., Riedel-Heller, S., Weyerer, S., Blank, W. A., et al. (2012). The influence of age, gender and socio-economic status on multimorbidity patterns in primary care.

First results from the multicare cohort study. *BMC health services research*, *12*(1), 1–15.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E., & Ungar, L. H. (2013). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0073791

Shneiderman, B. (1987). Designing the user interface strategies for effective human-computer interaction. *ACM SIGBIO Newsletter*, *9*(1). https://doi.org/10.1145/25065.950626

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4). https://doi.org/10.1016/j.ipm.2009.03.002

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1920484117

Stajner, S., & Yenikent, S. (2021). Why Is MBTI Personality Detection from Texts a Difficult Task? *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 3580–3589.

Stead, H., & Bibby, P. A. (2017). Personality, fear of missing out and problematic internet use and their relationship to subjective well-being. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2017.08.016

Stefan, F. (1996). *Mikroekonomiczne aspekty przepływu informacji między podmiotami rynkowymi*. Wyższa Szkoła Bankowa, Poznań.

Strelau, J. (1998). Individual differences in temperament: An international perspective. *Advances in psychological science*.

Strelau, J. (2019). Psychologia Temperamentu.

Strelau, J., & Zawadzki, B. (2012). Activity as a temperament trait. *Handbook of temperament.*

Strus, W., & Cieciuch, J. (2014). Poza Wielka Piatk – przeglad nowych modeli struktury osobowości. *Polskie Forum Psychologiczne*.

Strus, W., Cieciuch, J., & Rowiński, T. (2014). Polska adaptacja kwestionariusza IPIP-BFM-50 do pomiaru pięciu cech osobowości w ujęciu leksykalnym. *Roczniki Psychologiczne*, *17*(2), 327–346.

Sundar, S. S., & Marathe, S. S. (2010). Personalization versus Customization: the Importance of Agency, Privacy, and Power Usage. *Human Communication Research*, *36*(3), 298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x

Svendsen, G. B., Johnsen, J.-A. K., Almås-Sørensen, L., & Vittersø, J. (2013). Personality and technology acceptance: the influence of personality factors on the core constructs of the Technology Acceptance Model. *Behaviour & Information Technology*, *32*(4), 323–334.

Swaminathan, J. M., & Lee, H. L. (2003). Design for Postponement. https://doi.org/10.1016/S0927-0507(03)11005-5

Syrdal, D. S., Dautenhahn, K., Woods, S., Walters, M. L., & Koay, K. L. (2006). 'Doing the right thing wrong' - Personality and tolerance to uncomfortable robot approaches. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. https://doi.org/10.1109/ROMAN.2006.314415

Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: the five factor model of personality. *Journal of Experimental Psychology: Applied*, *17*(2), 71.

Teijeiro-Mosquera, L., Biel, J. I., Alba-Castro, J. L., & Gatica-Perez, D. (2015). What your face vlogs about: Expressions of emotion and big-five traits impressions in youtube. *IEEE Transactions on Affective Computing*, *6*(2). https://doi.org/10.1109/TAFFC.2014.2370044

Tett, R. P., Guterman, H. A., Bleier, A., & Murphy, P. J. (2000). Development and Content Validation of a "Hyperdimensional" Taxonomy of Managerial Competence. *Human Performance*. https://doi.org/10.1207/S15327043HUP1303_1

Tett, R. R., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1994). Meta-Analysis of Personality - Job Performance Relations : Replay to Ones, Mount, Barrick, and Hunter (1994). *Personnel Psychology*. https://doi.org/10.1111/j.1744-6570.1994.tb02415.x

Tomarchio, O., Calvagna, A., & di Modica, G. (2002). Virtual home environment for Multimedia services in 3rd generation networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2345*. https://doi.org/10.1007/3-540-47906-6_111

Trapnell, P. D., & Wiggins, J. S. (1990). Extension of the Interpersonal Adjective Scales to Include the Big Five Dimensions of Personality. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.59.4.781

Triandis, H. C., & Suh, E. M. (2002). Cultural influences on personality. *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.53.100901.135200

Trzcieliński, S. (2020). *Przedsiębiorstwo zwinne - Agile Enterprise*.

Tufekci, Z. (2018). Facebook's Surveillance Machine.

Tyagi, S., & Mittal, S. (2020). Sampling approaches for imbalanced data classification problem in machine learning. *Proceedings of ICRIC 2019* (pp. 209–221). Springer.

Van Hoek, R. I. (2001). Rediscovery of postponement a literature review and directions for research. *Journal of Operations Management*, *19*(2). https://doi.org/10.1016/S0272-6963(00)00057-7

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a framework for evaluation in design science research. *European journal of information systems*, *25*(1), 77–89.

Verhoeven, B., Daelemans, W., & Plank, B. (2016). Twisty: a multilingual twitter stylometry corpus for gender and personality profiling. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 1632–1637.

Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion Matrix-based Feature Selection. *MAICS*, *710*, 120–127.

Wagner, H. R., & Rogers, R. E. (1969). Max Weber's Ideal Type Theory. *Social Forces*, *48*(2). https://doi.org/10.2307/2575273

Wang, H., Yang, Y., Wang, H., & Chen, D. (2013). Soft-Voting Clustering Ensemble. In Z.-H. Zhou, F. Roli, and J. Kittler (Eds.), *Multiple Classifier Systems* (pp. 307–318). Springer Berlin Heidelberg.

Wang, P., Yan, Y., Si, Y., Zhu, G., Zhan, X., Wang, J., & Pan, R. (2020). Classification of Proactive Personality: Text Mining Based on Weibo Text and Short-Answer Questions Text. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.2995905

Wei, H., Zhang, F., Yuan, N. J., Cao, C., Fu, H., Xie, X., Rui, Y., & Ma, W. Y. (2017). Beyond the words: Predicting user personality from heterogeneous information. *WSDM 2017 - Proceedings of the 10th ACM International Conference on Web Search and Data Mining*. https://doi.org/10.1145/3018661.3018717

Wind, Y. (1978). Issues and advances in segmentation research. *Journal of marketing research*, *15*(3), 317–337.

Xu, R., Frey, R. M., & Ilic, A. (2016). Individual Differences and Mobile Service Adoption: An Empirical Analysis. *Proceedings - 2016 IEEE 2nd International Conference on Big Data Computing Service and Applications, BigDataService 2016*, 234–243. https://doi.org/10.1109/BigDataService.2016.15

Yildirim, C., & Correia, A. P. (2015). Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2015.02.059

Zawadzki, B. (2017). The location of personality disorders in the Circumplex of Personality Metatraits [Lokalizacja zaburzeń osobowości w kołowym Modelu Metacech Osobowości]. *Roczniki Psychologiczne*.

Zawadzki, B., Strelau, J., Szczepaniak, P., & Śliwińska, M. (1998). Inwentarz osobowości NEO-FFI Costy i McCrae. *Adaptacja polska. Podręcznik. Warszawa: Pracownia Testów Psychologicznych PTP*.

Zhang, J., Zheng, M., Nan, J., Hu, H., & Yu, N. (2020). A novel evaluation metric for deep learning-based side channel analysis and its extended application to imbalanced data. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 73–96.

Zhang, Q., Esterwood, C., Yang, J., & Robert, L. (2019). An Automated Vehicle (AV) like Me? The Impact of Personality Similarities and Differences between Humans and AVs. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3446005

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack Obama's books of 2019*. Profile books.

List of Tables

List of Figures

Appendices

Appendix A: Report From Pre-Research (Qualitative Part)

Appendix B: Psychometric Report From B5F5 and B5F6 Questionnaires (Big Five Personality Questionnaire Tool)

Appendix C: The brief report from data collecting fieldwork (used for preparing data-driven personality model

Appendix D: The list of all raw data collected by App *Dr Charakter*

A Report From Pre-Research (Qualitative Part)

A.1 Methodological Note

Survey was carried out in the first half of 2018, on 60 users of mobile phones, residents of Warsaw. In order to emphasize the diversity between people resulting from their personality, the study was conducted on a very homogeneous group in terms of demographics (the aim was to reduce confounding factors). The age has been limited to 20-29 years, separate groups for both genders (50% women, 50%men). Technologically advanced group, using many functionalities of mobile devices. The study was a multi-stage: filling of the personality questionnaire (Big 5), monthly observation of behaviors in social profiles and behaviours connected with smartphone usage. Then the in-depth structured interviews aimed at getting as much information as possible about behavior patterns were conducted. The results of these studies were behavioral metrics for each of 5 dimensions. The research delivered clear evidence that personality dimensions are good enough for discriminating users needs and expectations (see Tab. 1).

Moreover, respondents confirm that the services based on artificial intelligence are currently not adapted to them and this is the primary reason for their rejection or dissatisfaction with available services. The short summary in tabular data set shows (Tab. 1) how diverse the attitudes and expectations of mobile phone users towards virtual assistants are. Differences concern mainly the function of a virtual assistant and are a simple derivative of diversity of needs. At the same time, it can be seen that attitudes polarize with respect to personality dimensions. For example, people who are open to experience (high openness) expect non-standard content, have a high level of cognitive needs, and high need to explore (curiosity). In turn, people with low openness expect only a sense of comfort in a world that is well known to them (they like only what they know, they are afraid of unknown). For another dimension, individuals with a high level of conscientiousness use mainly the functionalities that help in the implementation of the need for control, which mainly manifests itself in the control of time and in scrupulous planning. On the other hand people with a low level of conscientiousness, who accept life in chaos and disorder, need only very basic control functionalities and will never be interested in the use of advanced calendar or notebook functions.

Based on the interviews and observation on Facebook, the factors that can affect

Table 1. Differentiation of needs related to the smartphone usage dependable on personality dimensions (Big 5) (source: own research)

Personality dimension	High level	Low level
E	Contact with people, facilitates face to face contact	Close-up with people - easy contact with loved ones (writing removes barriers that hinder f2f contacts)
O	Search tool, mini computer, searching for functionality, new technologies, saves time of exploration. Contact with others similar to them	Being up-to-date, quick information when I need it; pragmatism
C	Remembering, organising, acquiring information (control tool)	Contact with friends, check information (it's easier to grasp)
N	Phone helps in work, school, situations requiring quick response, key information (quick find)	There is no dominant need, they investigate many different situations: "serves to live"
A	Facilitating the acquisition of information, makes life easier (convenience,all in one)	"I use it when I have a need", mainly for the contact with the world

In qualitative study the IPIP Big 5 tool was used, and it has opposite scale of Stability called Neuroticism, Low Neurocitism is equivalent of High Emotional Stability and consequently High Neurocitism is equivalent of Low Emotional Stability

behaviour in the digital are identified:

- DEMOGRAPHY: Wealth (the amount of data to consume and the frequency of using WiFi); Life period (the beginning of university is the foundation of a student group on FB); Place of residence and distance to the place of work / study (moving between BTS).
- INFLUENCE OF OTHERS: Influence of a group of peers (for example by encouraging extreme behaviors on the internet), Fashion for specific behaviors (for example using Snapchat as event documentation); Education (e.g.the principle "do not pick up your phone from people that you do not know well").
- EVENTS IN LIFE: Serious life change, for example after the birth of a child, the user starts publishing photos of children; Disease / ailment (using the voice recorder to correct wrong diction); Events in social media > cautiously publishing posts after blocking an account or an account break-in.

Table 2. Different attitudes and expectations connected with virtual assistant (VA) dependable on personality dimensions (Big 5) (source: own research)

Personality dimension	High level	Low level
E	rejection of VA, but if adapted to the user expectations (e.g. no advertisements) can be considered	reject based on privacy (need for high protection), VA in role of invisible friend
O	the repeatability based on usage history is irritating but the idea of smart proactive VA searching all they need is attractive, they would be first to try it, their interest is changing and evolving	they expect to do something easier and faster - but still only when they need it (they initiate the action) - their interests is rather stable
C	the only group which like personalization, VA for them can be useful tool for controlling themselves „keep schedule & order"	They are aware of their own problems so they want the functionalities that helps them avoid errors due to lack of systematic or discretionary decision-making etc.
N	They appreciate the convenience of this solution (it limits forgetting, helps), but skeptical about data transfer and risk of "manipulation"	If they will have enough time for adaptation they can go with it (the value for them is an effectiveness (in real life))
A	They are afraid but if it became popular (recommended by friends) they will accept it	Their main motivation is a pragmatic needs (usefulness in everyday life).

In qualitative study the IPIP Big 5 tool was used, and it has opposite scale of Stability called Neuroticism, Low Neurocitism is equivalent of High Emotional Stability and consequently High Neurocitism is equivalent of Low Emotional Stability

- TRENDS: General consumer trends (eg smart-shopping and the tendency to limit spending on online services, including paid applications, more mobile content than TV content; Technological trends (e.g., displacing SMS as a limited form by Messenger).
- COMPETENCES: Competences in the digital world and the attitude of loved ones to using a smartphone (for example SMS to family vs Messenger to friends).
- CONTROL OVER CONTENT: Controlling content shared with moderators (possible exclusion from group of friends in SN or limiting display of contents); Awareness of the consequences of publishing content in the digital world (users know that profiles are observed by potential employers).

- SOCIAL MEDIA CULTURE: Proper ways of behaviour in digital world (recognised by culture) social media promote openness, presenting high self-confidence, happiness, ideal life.
- PURPOSE OF USING PHONE: The nature of using a mobile phone (business/private reasons) and the nature of the work (in one place or on business trips, contact with clients) can affect the readings of movement between the BTS or the analysis of the number of calls to specific phone numbers. Business profile on SM, activities indicated by need of business promotion and building the network (engaging people).

A.2 Key Findings

The study provides clear evidence that personality is an important determinant of behaviour in the digital world. Hence, in the development of this data, it is certainly possible to create valuable models. The resulting metrics for individual features, however, require more complicated processing than is originally assumed. Many behaviour are diagnostic only in context. The success depends on the presidency and ability to reproduce defined metrics based on data. It seems that the mere division of the study into groups that have consistent personalities allowed to interesting insights that would normally be omitted to be covered. Use of personality characteristics of users may have a key meaning in improving customer experience and customer satisfaction. When it comes to building a valuable personal assistant service, the key will be 3 out of 5 personality dimensions (according to the Big 5 model) :

- Openness High: do not accept algorithms based on user history.
- Extraversion High: due to the strong need to contact a living person.
- Conscientiousness High: strongly defined expectations must be a tool to help control somebody's life.

Other features may be significant, but they do not determine the overall design of the service. Neuroticism is an indication of the possible changing emotional moods of the user, while Agreeableness seems to be of great importance in communication with the client itself and not necessarily in the functionality of the service.

B Psychometric Report From B5F5 and B5F6 Questionnaires (Big Five Personality Questionnaire Tool)

B.1 Assumptions and hypotheses

The assumption made for the psychometric procedure concerns most of all the aim for which the tool is created (McCrae and Costa, 1987). The tool will be tested in the electronic version of the survey (online research), distributed via a website [www.szoprojekt.pl]. The IPIP-BFM-50 survey was also carried out in the electronic version, and its equivalence of measurement in different survey conditions (paper and pencil versus online survey) has been confirmed through Confirmatory Factor Analysis. It should be noted that the final tool will not be used for diagnostic purposes but research and classification of the trait levels (low, medium or high) and define the dimension dominant for a specific person. Before carrying out the research, the following research hypotheses have been formulated:

- H1. B5PF tool will have reliability satisfactory for scientific research. The reliability will be verified using Cronbach's alpha coefficients.
- H2. B5PF tool will have the external validity satisfactory for scientific research. The external validity will be verified through correlation analysis with an already existing and verified tool, which is IPIP-BFM-50 (Costa and McCrae, 1992). In addition, a comparative analysis of distributions for differentiation of personality traits according to gender and age obtained for five personality dimensions will be carried out.
- H3. B5PF tool will have a satisfactory theoretical validity. Its verification will be conducted using Confirmatory Factor Analysis.

B.2 Method

Generating Questionnaire Items. As a first step, 60 questions (12 for each dimension) have been formulated, based on the existing theoretical and domain knowledge and results of the qualitative research (Appendix A). The participants of the research were recruited using the IPIP-50 questionnaire. The individual interviews were carried out with 58 people representing personality profiles in which at least one trait was

Distribution of sex and age for whole sample (n=1128)

age intervals	total N=1128	men N=478	women N=650
18-29	0.27	0.28	0.27
30-39	0.32	0.30	0.33*
40-49	0.30	0.32*	0.28
50+	0.11	0.10	0.11

dominant (top or bottom 30% of the distribution). The test items are constructed as behavioural statements. For each dimension, six items for each extreme of a dimension were chosen using the method of competent judges. In the pool of 60 statements, none of the extremes for the five dimensions was over-represented. The judges (3 domain experts) chose the items based on matching between behavioural descriptions obtained from the qualitative research and the differentiation of behaviours between the items.

Research Fieldwork. The research was carried out on a group of volunteers. The website with the survey [www.szoprojekt.pl] was distributed via Internet communication (word of mouth), mainly through different sorts of social media, such as Plazza and FB. The research was conducted between 20th February 2019 and May 2019. There were 1333 surveys collected altogether, and 1131 of them were qualified for the research. The eliminated surveys were considered 'clicked' without consideration, as the time spent filling them in was too short to allow careful thinking. The participants were to complete two surveys: the first one contained 60 questions – proposals, among which the final items of B5PF items were to be chosen, and the second one contained 50 questions from IPIP-BFM-50 (Costa and McCrae, 1992; Zawadzki et al., 1998). The demography was limited to two questions about gender and age.

The obtained distribution does not diverge significantly from the distribution in the population aged 18-49, however, the number of participants aged 50+ is significantly lower. It results from the recruitment method of the research participants (a website, social media, corporate portal). Distribution of sex and age confirms that the data was obtained from a diversified group of people.

B.3 Results and Discussion

The following table contains descriptive statistics from both surveys B5PF-5 and IPIP-BFM-50. The distribution of results for each of the Big Five dimensions is similar to the

Table 3. Reliability (alfa-Cronbach) and descriptive statistics of questionnaire scales for both tools used (after scaling to a range of 0-100).

Dimension	Tool	M	SD	Skewness	Curtosis	alfa-Cronbach
Extraversion	B5PF-6	52.5	23.3	-0.239	-0.716	0.79
	IPIP-BFM-50	53.5	21.7	-0.131	-0.560	0.82-0.91
Agreeableness	B5PF-6	74.4	14.7	-0.868	0.993	0.60
	IPIP-BFM-50	72.9	14.8	-0.532	0.068	0.79-0.82
Conscientiousness	B5PF-6	61.9	19.9	-0.261	-0.363	0.62
	IPIP-BFM-50	65.9	17.4	-0.325	-0.364	0.75-0.83
Emotional Stability	B5PF-6	55.5	22.4	-0.163	-0.639	0.73
	IPIP-BFM-50	50.2	22.2	0.040	-0.590	0.86-0.90
Openness/ Intellect	B5PF-6	61.1	17.5	-0.267	-0.068	0.65
	IPIP-BFM-50	67.8	15.3	-0.226	-0.310	0.70-0.78

Table 4. Cronbach's Alfa reference for IPIP-BFM-50 given after (Costa and McCrae, 1992) as a range, the source text gives values separately for 8 independent studies (N from 304-2347), a total of 7015 people.
** This dimension has different names in tools: Openness in B5PF-6 and Intellect in IPIP-BFM-50.

Table 5. Cronbach's Alfa, Pearson's r index and minimum and maximum of discriminative power of the position for the tested Big 5 tool

	Cronbach's Alfa	r-Pearson with IPIP	min. power of item	max. power of item
E	0.79	0.85	0.37	0.66
A	0.60	0.55	0.26	0.50
O	0.62	0.64	0.28	0.41
S	0.73	0.82	0.35	0.57
C	0.65	0.77	0.27	0.45

normal distribution. The skewness and the kurtosis are in the range (-0.868 to 0.993) and they have the highest values for Agreeableness.

B.3.1 Measurement Reliability

Measurement reliability for psychometric tools is indicated by Cronbach's Alfa calculated separately for each of the five personality dimensions. The following table shows the primary indicators for both versions B5PF-5 and B5PF-6.

Cronbach's Alfa indexes above 0.7 can be considered high, and such they are for the extraversion and neuroticism (see and compare with [4]). Slightly lower measures were obtained by other scales, i.e. Conscientiousness, Openness and Agreeableness. It was possible to configure items with higher Cronbach's alfa coefficients but with simultaneous deterioration of external validity (Pearson's r from IPIP). From the point of view

Table 6. Cronbach's alfa, Pearson's r index and discriminatory power of the position for the tested B5PF

r-Pearson z IPIP	E	A	O	S	C
B5PF-6	0.85	0.55	0.64	0.82	0.77

of the tool's purpose, it was considered that it is best to balance both parameters and items was selected to obtain an acceptable level of reliability with the highest possible Accuracy.

B.3.2 External Accuracy

The Extraversion and Stability scales have the highest accuracy in both tools. Conscientiousness and Openness scales also have good accuracy. Lower accuracy for Openness may result from its different scope than in the IPIP tool, where this scale is defined as Intellect. In B5PF scales, Openness is defined more broadly than Intellect in IPIP. Because of business goals, we were more interested in measuring Openness to experience. This feature may have crucial importance in Openness to new unknown technologies, services' functions. In turn, Intellect in IPIP is defined as a tendency to consider and think more closely to cognitive involvement. This difference in definition range may result in lower accuracy. The r-Pearson indices for all scales are presented in Tab4.

The lowest external validity index was obtained for the Agreeableness scale. This scale is the weakest in B5PF-6. Our observations and conversations with people completing surveys show that this scale is most "contaminated" by the variable of need for social approval. This hypothesis will be verified in subsequent stages of the study (for the main study, a question about the need for social approval was added to the list of explanatory variables.

B.3.3 Confirmatory Factor Analysis as a confirmation of theoretical validity

Due to the way the tool is used and testing its accuracy using an external tool, there is no need to test the theoretical validity, which is essential when the tool is designed according to theoretical strategy (for checking the strength of the theoretical model when it testing external validity is not possible). Only for cognitive purposes and understanding the relationship between scales and additional position power verification, confirmatory factor analysis was conducted by the Amos 21 program (SPSS package).

Table 7. Table with different model fit indicators for both B5PF scales.

	Chi square - data sensitive test (n>200)		error of approximation to an ideal population (better for CFA) ideal is lower than 0.8	Model Fit Index (ideal is 0)	Model Fit Index better for exploitative FA (ideal is 1)		Goodness of Fit (ideal is 1)
Whole model	p	CMIN/DF	RMSEA	SRMR	NFI	CFI	GFI
B5PF-5	p<.0005	5.9	0.07	.07	.74	.78	.89
B5PF-6	p<.0005	6.6	.07	.08	.66	.70	.85
B5PF-5_without outliers	p<.0005	5.6	.06	.07	.75	.79	.89
Separate scales							
BSPF-5 - E	p<.0005	22.9	.13	.06	.93	.93	.96
BSPF-6 - E	p<.0005	17.6	.12	.05	.92	.92	.95
BSPF-5 - A	p=.108	1.8	.02	.01	.98	.99	.997
BSPF-6 - A	p=.005	2.6	.03	.02	.97	.98	.99
BSPF-5 - C	p<.0005	28.6	.15	.07	.82	.83	.95
BSPF-6 - C	p<.0005	29.2	.15	.08	.73	.73	.92
BSPF-5 - S	p<.0005	5.4	.06	.02	.98	.98	.99
BSPF-6 - S	p<.0005	29.2	.15	.08	.81	.82	.93
BSPF-5 - O	p<.0005	14.8	.11	.05	.86	.87	.97
BSPF-6 - O	p<.0005	1.8	.09	.05	.85	.87	.97

The tested model is shown in figure Fig3. Estimating the model's fit to the data was based on RMSEA, CFI and SRMR indicators. A value below 0.08 is widely accepted as acceptable RMSEA and SRMR. CFI should be above 0.9 (Hu and Bentler, 1999; Marsh et al., 2004).

As can be seen from the table, the condition for RMSEA is met for both scales. The whole 5-factor model has a better fit to the data than models for individual sub-scales. Unfortunately, CFI does not meet the conditions recommended by Hu and Bentler (1999) and Marsh et al. (2004) (value higher than 0.8). However, according to Rigdon (1996), CFI and RMSEA are indicators complementary to each other. CFI assumes that the base model (0) against which the estimate is made is correct. Therefore CFI is recommended for exploratory studies on small samples (e.g. N = 100). In contrast, RMSEA seems to work better for the confirmation model on large samples, which corresponds to the presented study.

B.4 Inter-correlations between the scales

The table below contains the inter-correlation coefficients between dimensions for the B5PF-5, B5PF-6 scale and the IPIP scale on the same group of subjects tested as a reference. Openness and Extraversion for both B5PF versions are the most correlated.

B.4.1 Variety of features in terms of gender and age

The following regularities were found based on survey B5PF5. The average level of extraversion, agreeableness, diligence and neuroticism is significantly higher among women than men. In turn, emotional stability is significantly higher among men.

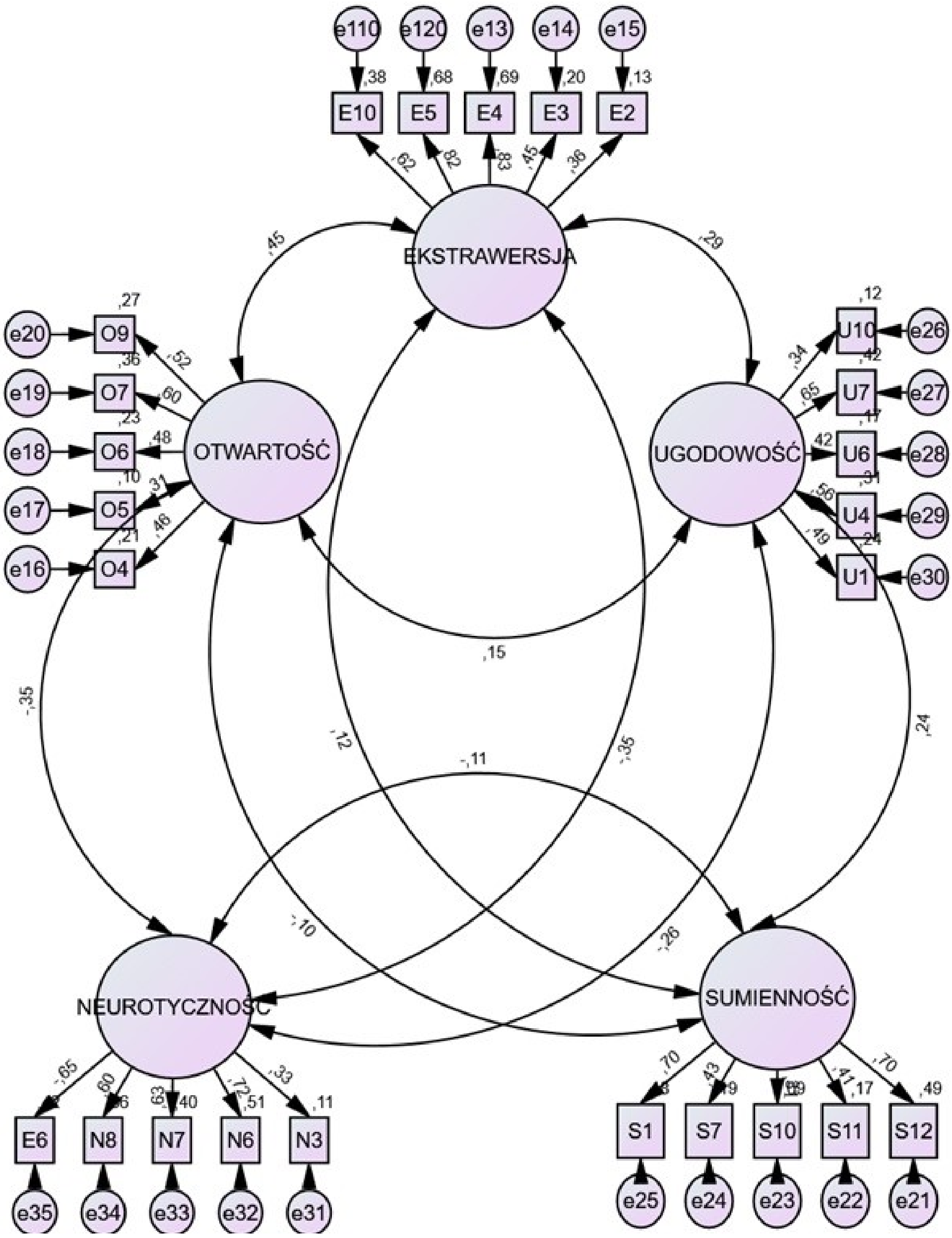


Figure 1. Model tested in CFA (Ekstrawersja means Extraversion, Otwartość means Openness, Ugodowość is Agreeableness, Neurotyczność is Emotional Stability and Sumienność is Conscientiousness)

(Source: own work)

Table 8. Table with internal correlations between the factors for B5PF-5 (top) and B5PF-6 (bottom).

	E	A	C	E	O
E		.209	.088	.228	.340
A	.148		.150	.226	.122
C	.148	.166		.153	-.025
S	.156	.218	.206		.199
O	.396	.087	.054	.191	

Correlation coefficients of personality dimensions according to B5PF_5 with age: extroversion (r = 0.098) agreeableness (r = 0.096), conscientiousness (r = 0.17), emotional stability (r = 0.12), openness to experience (-0.001). It is worth noting that stability increases significantly after 29 years of age, and then the diversity decreases. Conscientiousness and extraversion increase gradually, meaning both dimensions are subject to some degree of learning and result from adaptation to the environment.

B.5 Summary and Conclusions

Considering the above analyzes and statistics, it can be concluded that both versions of the questionnaire for the study of 5 Great personality factors (Big 5) meet the basic criteria for psychometric scales for scientific purposes. Satisfactory parameters were achieved for 4 out of 5 factors, the weakest though still acceptable is Agreeableness. Both 25 and 30 question versions can be used to classify and rank dimensions to determine the dominant personality profile features of clients. If you have more empirical material, you can always repeat the above analyzes to determine the current psychometric parameters of the tool. Repetition of analyzes is particularly recommended for normalizing results. The research hypothesis to be developed on more numerous samples is certainly to verify whether the worse parameters for the Agreeableness factor are not the result of a high skewness of the distribution of raw results.

C The complementary information from data collecting fieldwork (used for preparing data-driven personality model

The data on which the UISPP model was built was collected in two ways. Data from about 1,800 people (67.5% of the data set) were obtained through a paid recruitment from the research panel of ARC Rynek i Opinia (It is a recognized company that has been operating in Poland for many years, carrying out marketing research). The remaining number of interviews (866) was collected thanks to crowdsourcing (similar to the collection of the sample for the psychometric test - Appendix B). As for crowdsourcing, it was mainly implemented through social media, both of a business nature (Plazza - which is an internal social portal for Orange Polska employees) and external, such as Facebook, using private social networks connections using the snowball effect. It should also be emphasized, that persuading people to install an application containing 5 legal consents for processing, saving and accessing data is not trivial task. The graph below shows the flow of data in the subsequent months of 2020. The original assumptions for the project assumed collecting data from 5,000 people. Unfortunately, it turned out to be impossible within six months, which is why a paid recruitment for the study was decided.

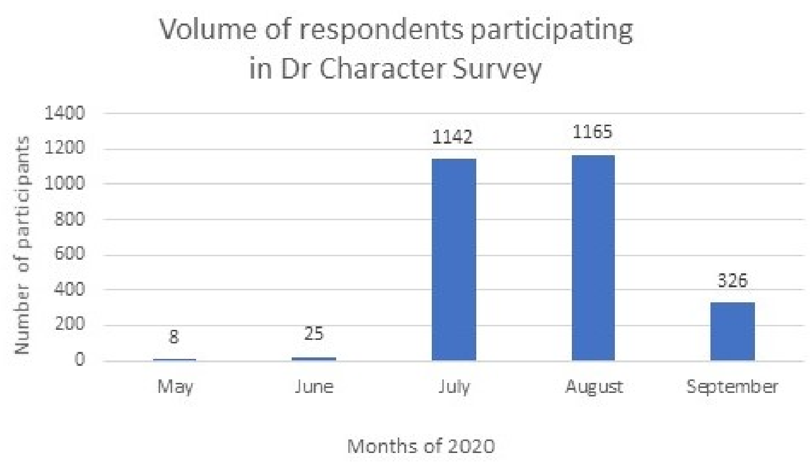


Figure 2. Volume of respondents participating in Dr Character Survey in months of 2020

(Source: own work)

Finally, the application was installed and research conditions (consents) were accepted by 2,684 and 2,666 people decided to participate in the study, including 1,303 men and 1,364 women. As far as the nature of the obtained sample is concerned from the research point of view, we can speak of a purpose-quota sample in this case. The sample selection was controlled in terms of sex and age (two groups up to 30 years and over 30 years). Just taking part in the study required specific numerical skills, understanding commands, and consent to share the animated phone data for research purposes. According to data from a research company that recruited people for the study from its panel, the response rate was below 10% of invitations sent. The most appropriate sample description is technologically advanced people to the extent that allows them to perform the test independently and, at the same time, people who trust the exhibitor.

C.1 Mobile Application

The Dr Charakter android application was written by front-end and back-end developers and designers at Orange Labs Polska based on defined requirements. The entire project team consisted of 5 people. The application was a self-realizable study that, without the need for assistance, conducted the test subject through successive stages of the study, from explaining the purpose of the study, collecting the necessary consents and explaining what the data is for, and through an internal analytical module calculating statistics from the data on the phone. Fully anonymous data was sent in several packages to the laboratory server. The test itself lasted from 3 to 10 minutes. converting and sending data up to several hours. Unfortunately, the processing of photo data on the phone turned out to be problematic and for about 30% of people it was not possible to obtain a complete set of data from the photo analysis. This made this type of data excluded from analysis. It was concluded that the final UISPP model cannot be burdened with such a high risk of failure.

C.2 Application Screen Script

Original script used in *Dr Charakter* application was in Polish. Here below is translated version. Based on psychometric practice the diagnostic questions for Big 5 profile

(B5PF5) will not be published.

Intro from the application and consent What is Dr Charakter? The purpose of Dr Charakter is to obtain only the ANONYMOUS data needed for laboratory research into differences in human behavior.

If you honestly answer 25 questions, you will find out what kind of personality you have in the 5 dimensions. Follow the instructions and you will learn interesting things about yourself! Some may surprise you.

What's waiting for you? Participation in the survey is completely voluntary and anonymous, for the analysis we need to complete a survey, information about gender and age, and calculated statistics illustrating how you use your phone. Nothing more.

Finally, we will show your score on the 5 personality dimensions and your stats. The study is intended for adults. Before you start the study, we will ask you to accept the consents required by law.

The application will ask Android for access to the following data: *Contacts* To be able to count the number of people you contact

The application will ask Android for access to the following data: *Call history* To check if how long and often you talk

The application will ask Android for access to the following data: *Multimedia* To count how many photos you took and what kind

The application will ask Android for access to the following data: *SMS* To check if you write or call more often.

SCREEN with consent [] I agree that Orange Polska, in order to enable me to participate in the study, process the above-mentioned data on my device and send the received report to its IT system.

Providing data is voluntary, it is not an obligation resulting from the provisions or the contract.

I know that until the data is anonymized, I can withdraw my consent in the application (it will not affect the lawfulness of the use of my data before the consent is withdrawn).

Anonymous research results will be available to our partners, research and scientific institutions.

The data administrator is Orange Polska. More information in the Application /

Research Privacy Policy [...]

The result of the personality test is informative and does not constitute a psychological diagnosis within the meaning of the law. To obtain such a diagnosis, you should contact a psychologist.

The application, test, content of questions is protected by copyright.

Pre-survey instruction There are 25 behavioral statements in front of you. Decide about each of them to what extent it also suits you. Each answer is correct when it relates to your most common behaviors.

[Then the 25 statements were presented. One on the screen at a time. With visualised Likert scale. After answering 25 questions and some additional about gender, age and the and style of using the phone (private, business, mixed]

In the end research participant received its Big 5 profile 9 with few screens of statistics calculated from phone usage variables.

Table 9. The description of traits exposed to the Dr Charakter research participants.

Personality Trait	High level	Low Level
Extraversion	You are of an energetic, talkative, friendly and sociable person. You have a lot of energy and share it enthusiastically with others. You like to be the centre of attention, and you feel best among people. You are firm and often become the natural leader of the group.	You are a quiet, shy, reserved and silent person. You value true friendships and need a lot of time to open up to someone. People tire you, and you feel best in your own company.
Agreeableness	You are a kind, polite and cordial person. You never want to offend anyone, and you trust others. You enjoy helping others. You are generous in praising others and forgiving easily.	You are relentless in a person, you stick to your point of view, and therefore you often argue. Truth and right are more important to you than courtesy. You do not trust people, you see them as enemies rather than allies, and you have a hard time forgiving the mistakes of others.
Conscient-iousness	You are a well-organized, scrupulous and reliable person, responsible and prudent. You follow the rules, and the decisions you make are deliberate.	You are a person who functions well in chaos, paying no attention to order, plans and time frames. You act spontaneously and sometimes recklessly; you are not looking for perfection in life.
Emotional Stability	You are a stable, calm person, not guided by emotions, satisfied with life, usually relaxed and at ease. You are a controlled person, unshakable and detached from what is happening around you.	You are a restless, tense and nervous person. You often feel irritated by your surroundings and explode quickly, which you usually regret afterwards. You are easily annoyed by your surroundings, and you are often in a bad mood and have poor well-being.
Openness for experience	You are a person who is curious about the world, inventive, creative and with great imagination, for whom aesthetics and values (ideas) are essential, with extensive interests. You are often guided by your intuition and do not tolerate boredom in life.	You are a person with constant preferences, disliking changes and carefully accepting new products. You value ordinary, uncomplicated and straightforward things. You only like what you know well, relying on your knowledge and experience.

D The Lists of Variables (UISSP model)

The first table below (10 consists of dictionary of raw variable list used for creating model (gathered from Dr Character). The pictures and exifs variables are not presented because they were excluded, on the early stage. Time of processing it was not acceptably long. Three categories of data (Android usage, Application info and Application statistics) are not presented in the table because they are listed on Section 3.5.2.

The second table 11 presents the actual complete list of 261 features used for creating UISPP selected based on Mutual Information score. This list was reduced until achieving final version of UISPP model described in Chapter 4.

Table 10. Raw Variable List - data from Dr Character App.

Row Variable Name	**Variable description**
Call logs data variables:	
timestamp	Date when call was performed.
number_id	anonymised phone number of secondary user for call.
type	Type of logged call (incoming, outgoing, missed, blocked, rejected)
duration	Duration of calls in seconds
SMS logs data variables:	
timestamp	Date of sms was sent/received
number_id	Anonymised phone number of secondary user for call or sms.
type	Sms type: incoming, outgoing.
number_type (for sms)	Types: short, mobile, unknown
average_word_count_sent	Average word count from sms send by user.
average_word_length_sent	Average word length in sms send by user.
average_sign_count_sent	Average count of punctuation marks in sms send by user.
average_emoticon_count_sent	Average emoticon count in sms send by user.
emoticon_sms_portion_sent	Portion of sms send by user containing at least one emoticon.
average_word_count_rec	Average word count from sms received by user.
average_word_length_rec	Average word length in sms received by user.
average_sign_count_rec	Average count of punctuation marks in sms received by user.
average_emoticon_count_rec	Average emoticon count in sms received by user.
emoticon_sms_portion_rec	Portion of sms received by user containing at least one emoticon.
mms_sent	Count of mms send by user. 12 columns representing last 12 months
Call and SMS logs statistics:	
calls_number_all	Weekly mean value of number of outgoing and incoming calls
calls_number_inc	Weekly mean value of number of incoming calls
calls_number_out	Weekly mean value of number of outgoing calls
text_messages_number_all	Weekly mean value of number of outgoing and incoming calls and text messages
text_messages_number_inc	Weekly mean value of number of incoming calls and text messages
text_messages_number_out	Weekly mean value of number of outgoing calls and text messages
calls_text_messages_number_all	Weekly mean value of number of outgoing and incoming calls and text messages
calls_text_messages_number_inc	Weekly mean value of number of incoming calls and text messages
calls_text_messages_number_out	Weekly mean value of number of outgoing calls and text messages
duration_all	Weekly mean value of duration of outgoing and incoming calls
duration_inc	Weekly mean value of duration of incoming calls
duration_out	Weekly mean value of duration of outgoing calls
calls_contacts_all	Weekly mean value of distinct contacts based on outgoing and incoming calls
calls_contacts_inc	Weekly mean value of distinct contacts based on incoming calls
calls_contacts_out	Weekly mean value of distinct contacts based on outgoing calls
text_messages_contacts_all	Weekly mean value of distinct contacts based on outgoing and incoming text messages
text_messages_contacts_inc	Weekly mean value of distinct contacts based on incoming text messages
text_messages_contacts_out	Weekly mean value of distinct contacts based on outgoing text messages
calls_text_messages_contacts_all	Weekly mean value of distinct contacts based on outgoing and incoming calls and text messages
calls_text_messages_contacts_inc	Weekly mean value of distinct contacts based on incoming calls and text messages

Table 10 continued from previous page

Row Variable Name	Variable description
calls_text_messages_contacts_out	Weekly mean value of distinct contacts based on outgoing calls and text messages
calls_contacts_business_time_inc	Weekly mean value of distinct contacts based on incoming calls business time
calls_contacts_business_time_all	Weekly mean value of distinct contacts based on outgoing and incoming calls during business time
calls_contacts_business_time_out	Weekly mean value of distinct contacts based on outgoing calls business time
text_messages_contacts_business_time_all	Weekly mean value of distinct contacts based on outgoing and incoming text messages during business time
text_messages_contacts_business_time_inc	Weekly mean value of distinct contacts based on incoming text messages during business time
text_messages_contacts_business_time_out	Weekly mean value of distinct contacts based on outgoing text messages during business time
calls_text_messages_contacts_business_-time_all	Weekly mean value of distinct contacts based on outgoing and incoming text messages and calls during business time
calls_text_messages_contacts_business_-time_inc	Weekly mean value of distinct contacts based on incoming text messages and calls during business time
calls_text_messages_contacts_business_-time_out	Weekly mean value of distinct contacts based on outgoing text messages and calls during business time
calls_contacts_free_time_all	Weekly mean value of distinct contacts based on outgoing and incoming calls during free time
calls_contacts_free_time_inc	Weekly mean value of distinct contacts based on incoming calls free time
calls_contacts_free_time_out	Weekly mean value of distinct contacts based on outgoing calls free time
text_messages_contacts_free_time_all	Weekly mean value of distinct contacts based on outgoing and incoming text messages during free time
text_messages_contacts_free_time_inc	Weekly mean value of distinct contacts based on incoming text messages free time
text_messages_contacts_free_time_out	Weekly mean value of distinct contacts based on outgoing text messages free time
calls_text_messages_contacts_free_time_all	Weekly mean value of distinct contacts based on outgoing and incoming text messages and calls during free time
calls_text_messages_contacts_free_time_inc	Weekly mean value of distinct contacts based on incoming text messages and calls during free time
calls_text_messages_contacts_free_time_out	Weekly mean value of distinct contacts based on outgoing text messages and calls during free time
calls_avg_diff	Mean inter-event time calculated for outgoing calls
calls_diff_stdev	Standard deviation of inter-event time calculated for outgoing calls
text_messages_avg_diff	Mean inter-event time calculated for outgoing text messages
text_messages_diff_stdev	Standard deviation of inter-event time calculated for outgoing text messages
calls_text_messages_avg_diff	Mean inter-event time calculated for outgoing calls and text messages
calls_text_messages_diff_stdev	Standard deviation of inter-event time calculated for outgoing calls and text messages
calls_contacts_entrophy_all	Contacts entrophy calculated for outgoing and incoming calls
calls_contacts_entrophy_inc	Contacts entrophy calculated for incoming calls
calls_contacts_entrophy_out	Contacts entrophy calculated for outgoing calls
text_messages_contacts_entrophy_all	Contacts entrophy calculated for outgoing and incoming text messages
text_messages_contacts_entrophy_inc	Contacts entrophy calculated for incoming text messages
text_messages_contacts_entrophy_out	Contacts entrophy calculated for outgoing text messages
calls_text_messages_contacts_entrophy_all	Contacts entrophy calculated for outgoing and incoming calls and text messages
calls_text_messages_contacts_entrophy_inc	Contacts entrophy calculated for incoming calls and text messages
calls_text_messages_contacts_entrophy_out	Contacts entrophy calculated for outgoing calls and text messages
calls_12_hour_response_ratio	Ratio of missed calls answered within 12 hour to all incoming calls
text_messages_12_hour_response_ratio	Ratio of text messages answered within 12 hour to all incoming calls
calls_text_messages_12_hour_response_ratio	Ratio of missed calls and text messages answered within 12 hour to all incoming calls
calls_12_hour_response_time	Mean response time for missed calls answered within 12 hours
text_messages_12_hour_response_time	Mean response time for text messages answered within 12 hours
calls_text_messages_12_hour_response_time	Mean response time for missed calls and text messages answered within 12 hours
calls_1_hour_response_ratio	Ratio of missed calls answered within 1 hour to all incoming calls
text_messages_1_hour_response_ratio	Ratio of text messages answered within 1 hour to all incoming calls
calls_text_messages_1_hour_response_ratio	Ratio of missed calls and text messages answered within 1 hour to all incoming calls
calls_1_hour_response_time	Mean response time for missed calls answered within 1 hour
text_messages_1_hour_response_time	Mean response time for text messages answered within 1 hour
calls_text_messages_1_hour_response_time	Mean response time for missed calls and text messages answered within 1 hour
calls_regularity	Regularity at which user is making calls
text_messages_regularity	Regularity at which user is sending text messages
calls_text_messages_regularity	Regularity at which user is making calls or sending text messages
calls_ar_1, (4, 8, 12, 24)	1st, 4th, 8th, 12th, 24th auto-regression time series model coefficients. Model constructed for outgoing calls
text_messages_ar_1 (4, 8, 12, 24)	1st, 4th, 8th, 12th, 24th auto-regression time series model coefficients. Model constructed for outgoing text messages
calls_text_messages_ar_1 (4, 8, 12, 24)	1st, 4th, 8th, 12th, 24th auto-regression time series model coefficients. Model constructed for outgoing calls and text messages
Contact List Variables:	
contacts_pager	Number of contacts with pager number
contacts_personal_number	Number of contacts with personal number
contacts_premium_rate	Number of contacts premium numbers
contacts_shared_cost	Number of contacts with shared cost
contacts_uan	Number of contact with UAN
contacts_unknown	Number of contacts with unknown number
contacts_ussd	Number of ussd numbers
contacts_voicemail	Number of voicemail numbers
contacts_voip	Number of VOP numbers

Table 10 continued from previous page

Row Variable Name	Variable description
contacts_with_duplicated_numbers	Number of contacts with duplicated numbers
contacts_fixed_or_mobile	Number of fixed_or_mobile
emails	Number of contacts with at least one email address
contacts	Number of user contacts
contacts_TYPE_HOME	Number of contacts with at least one home number bu user.
contacts_TYPE_MOBILE	Number of contacts with at least one mobile number.
contacts_TYPE_WORK	Number of contacts with at least one work number.
contacts_TYPE_UNIDENTIFIED	Number of contacts with at least one unidentified number.
contacts_address	Number of contacts with at least one address defined by user .
contacts_photo	Number of contacts with a photo.
contacts_ICE	Number of contacts with Group containing word ICE.
contact_one_week	Number of contacts which user called last week divided by total number of user contacts.
contact_two_weeks	Number of contacts which user called last 2 weeks divided by total number of user contacts.
contact_one_month	Number of contacts which user called last month divided by total number of user contacts.
contact_three_months	Number of contacts which user called last three months divided by total number of user contacts.
contact_six_months	Number of contacts which user called last six months divided by total number of user contacts.
contacts_family	Number of contacts that in name contains phrase linked with family dictionary set in application code
contacts_mobile	Number of contacts which phone number is recognized as mobile number 48AAXXXXXXX with Polish mobile prefixes
contacts_fixed	Number of contacts which phone number is recognized as fixed number 48AAXXXXXXX with Polish regional prefixes .
contacts_short	Number of contacts which phone number is shorter than 9 digits
contacts_foreign	Number of contacts other than previous groups (longer than 9 or no polish prefixes)
contacts_family_numbers	Ids of contacts that in a name has a phrase from family dictionary.
Android System Info variables :	
is_device_secured	Yes or no
is_keyguard_secure	Yes or no
lock_pattern_enabled	Yes or no
int mcc	IMSI MCC (Mobile Country Code)
int mnc	IMSI MNC (Mobile Network Code)
int density_dpi	The target screen density being rendered to, corresponding to density resource qualifier.
int screen_layout	Bit mask of overall layout of the screen.
String locale_tags	Current user preference for the locale
String locale	Current user preference for the locale
int color_mode	Bit mask of color capabilities of the screen.
float font_scale	Current user preference for the scaling factor for fonts
int hard_keyboard_hidden	A flag indicating whether the hard keyboard has been hidden.
int keyboard	The kind of keyboard attached to the device.
int keyboard_hidden	A flag indicating whether any keyboard is available.
int navigation	The kind of navigation method available on the device.
int navigation_hidden	A flag indicating whether any 5-way or DPAD navigation available.
int orientation	Overall orientation of the screen.
int ui_mode	Bit mask of the ui mode. Currently there are two fields.
int version_sdk	The SDK version of the software currently running on this hardware device.
String version_base_os	The base OS build the product is based on.
String version_release	The user-visible version string.
float battery_level	Battery Level in range 0-1
boolean battery_is_charging	Flag if battery is being charged
boolean battery_is_charging_usb	True if charging is via USB
boolean battery_is_charging_ac	True if charging is via AC power adapter
boolean battery_saver_mode_enabled	Returns true if the device is currently in power save mode.
Int rotation_enabled	Control whether the accelerometer will be used to change screen orientation.
String ringer_mode	Returns “our own strings for “RINGER_MODE_NORMAL” “RINGER_MODE_-SILENT” “RINGER_MODE_VIBRATE”
String nfc	NFC status: NOT AVAILABLE, NOT ENABLED or ENABLED
Int mobile_data_enabled	flag if mobile data is enabled
long block_size	The size, in bytes, of a block on the file system.
long total_size	The total number of blocks on the file system.
long available_size	The number of blocks that are free on the file system and available to applications.
long free_size	The total number of blocks that are free on the file system, including reserved blocks (that are not available to normal applications).
long total_bytes	The total number of bytes supported by the file system.
long block_size_sd	As above but for potential SD card that might be emulated only
long total_size_sd	As above but for potential SD card that might be emulated only
long available_size_sd	As above but for potential SD card that might be emulated only
long free_size_sd	As above but for potential SD card that might be emulated only
long total_bytes_sd	As above but for potential SD card that might be emulated only
boolean external_emulated	If external memory is emulated on local memory or not
int alarm_alert	Persistent store for the system-wide default alarm alert.
int bluetooth_discoverability	Determines whether remote devices may discover and/or connect to this device.
bluetooth_discoverability_timeout	Bluetooth discoverability timeout.
String date_format	Date format string mm/dd/yyyy dd/mm/yyyy yyyy/mm/dd
int dtmf_tone_type_when_dialing	CDMA only settings DTMF tone type played by the dialer when dialing.
int dtmf_tone_when_dialing	Whether the audible DTMF tones are played by the dialer when dialing.
int end_button_behavior	What happens when the user presses the end call button if they’re not on a call.
int haptic_feedback_enabled	Whether haptic feedback (Vibrate on tap) is enabled.
int mode_ringer_streams_affected	Determines which streams are affected by ringer and zen mode changes.

Table 10 continued from previous page

Row Variable Name	Variable description
int mute_streams_affected	Determines which streams are affected by mute.
int screen_brightness	The screen back-light brightness between 0 and 255.
int screen_brightness_mode	Control whether to enable automatic brightness mode.
int screen_off_timeout	The amount of time in milliseconds before the device goes to sleep or begins to dream after a period of inactivity.
int show_gtalk_service_status	Whether Gtalk service is enabled
int sound_effects_enabled	Whether the sounds effects (key clicks, lid open ...) are enabled.
int text_auto_caps	Setting to enable Auto Caps in text editors.
int text_auto_punctuate	Setting to enable Auto Punctuate in text editors.
int auto_replace	Setting to enable Auto Replace (AutoText) in text editors.
int text_show_password	Setting to showing password characters in text editors.
int time_12_24	Display times as 12 or 24 hours 12 24
int user_rotation	Default screen rotation when no other policy applies.
int vibrate_when_ringing	Whether the phone vibrates when it is ringing due to an incoming call.
int adb_enabled	Whether ADB is enabled.
int airplane_mode_on	Whether Airplane Mode is on.
String airplane_mode_radios	A comma separated list of radios that need to be disabled when airplane mode is on. This overrides WIFI_ON and BLUETOOTH_ON, if Wi-Fi and bluetooth are included in the comma separated list.
int always_finish_activities	If not 0, the activity manager will aggressively finish activities and processes as soon as they are no longer needed. If 0, the normal extended lifetime is used.
int animator_duration_scale	Scaling factor for Animator-based animations. This affects both the start delay and duration of all such animations. Setting to 0 will cause animations to end immediately. The default value is 1.
int auto_time	Value to specify if the user prefers the date, time and time zone to be automatically fetched from the network (NITZ). 1=yes, 0=no
int auto_time_zone	Value to specify if the user prefers the time zone to be automatically fetched from the network (NITZ). 1=yes, 0=no
int bluetooth_on	Whether bluetooth is enabled/disabled 0=disabled. 1=enabled.
int boot_count	Boot count since the device starts running API level 24.
int contact_metadata_sync_enabled	Whether to enable contacts metadata syncing or not The value 1 - enable, 0 - disable
int data_roaming	Whether or not data roaming is enabled. (0 = false, 1 = true)
int development_settings_enabled	Whether user has enabled development settings.
int device_provisioned	Whether the device has been provisioned (0 = false, 1 = true).
String http_proxy	Host name and port for global http proxy.
int stay_on_while_plugged_in	Whether we keep the device on while the device is plugged in.
int transition_animation_scale	Scaling factor for activity transition animations.
int usb_mass_storage_enabled	USB Mass Storage Enabled
int use_google_mail	If this setting is set (to anything), then all references to Gmail on the device must change to Google Mail.
int wifi_on	Whether the Wi-Fi should be on. Only the Wi-Fi service should touch this.
int wifi_sleep_policy	The policy of Wi-Fi should go to sleep
int wifi_watchdog_on	Whether the Wi-Fi watchdog is enabled.
int window_animation_scale	Scaling factor for normal window animations. Setting to 0 will disable window animations.
accessibility_display_inversion_enabled	Setting that specifies whether display color inversion is enabled.
int accessibility_enabled	If accessibility is enabled.
String allowed_geolocation_origins	Origins for which browsers should allow geolocation by default.
String android_id	A 64-bit number (expressed as a hexadecimal string)
String enabled_accessibility_services	List of the enabled accessibility providers.
String enabled_input_methods	List of input methods that are currently enabled.
int touch_exploration_enabled	If touch exploration is enabled.
int tts_default_pitch	Default text-to-speech engine pitch. 100 = 1x
int tts_default_rate	Default text-to-speech engine speech rate. 100 = 1x
String tts_default_synth	Default text-to-speech engine.
String tts_enabled_plugins	Space delimited list of plugin packages that are enabled.
String actual_default_ringtone_ur	A Uri pointing to the default sound for the sound type.

Table 11. Complete list of 261 features used for creating UISPP selected based on Mutual Information score. This list was reduced until achieving final version of UISPP model.

feature	data category	E (MI)	A (MI)	O (MI)	S (MI)	C (MI)
accessibility_display_inversion_enabled	phone settings	0	1	1	0	0
accessibility_enabled	phone settings	1	1	0	1	1
adb_enabled	phone settings	1	0	0	0	0
airplane_mode_on	phone settings	1	0	1	1	1
all_apps	aplication list	0	1	1	1	0
always_finish_activities	phone settings	1	0	0	0	0
animator_duration_scale	phone settings	0	1	0	0	0
apps_from_GooglePlayStore	aplication list	1	1	1	1	1
auto_replace	phone settings	0	0	1	1	1
auto_time	phone settings	0	1	0	1	0
auto_time_zone	phone settings	0	0	1	1	1
available_size	phone settings	0	1	1	1	1
available_size_sd	phone settings	0	1	0	0	0
battery_is_charging	phone settings	0	1	1	0	0
battery_is_charging_ac	phone settings	0	0	1	1	1
battery_is_charging_usb	phone settings	1	0	0	1	0

feature	data category	E (MI)	A (MI)	O (MI)	S (MI)	C (MI)
battery_level	phone settings	1	1	1	1	1
battery_saver_mode_enabled	phone settings	0	0	0	1	0
block_size_sd	phone settings	1	1	1	0	0
bluetooth_discoverability	phone settings	1	1	1	0	1
bluetooth_discoverability_timeout	phone settings	1	0	1	1	1
bluetooth_on	phone settings	1	0	0	1	0
boot_count	phone settings	0	0	1	0	0
category_is_Action_ratio	aplication list	1	0	1	0	0
category_is_Action_sum	aplication list	1	1	0	0	0
category_is_Adventure_ratio	aplication list	0	0	0	0	1
category_is_Adventure_sum	aplication list	1	0	0	1	1
category_is_Arcade_ratio	aplication list	0	0	1	1	0
category_is_Arcade_sum	aplication list	0	0	0	1	1
category_is_Art & Design_ratio	aplication list	0	0	0	1	0
category_is_Art & Design_sum	aplication list	0	1	0	0	0
category_is_Auto & Vehicles_ratio	aplication list	1	0	0	1	1
category_is_Auto & Vehicles_sum	aplication list	0	0	0	1	1
category_is_Beauty_ratio	aplication list	0	0	0	1	1
category_is_Beauty_sum	aplication list	0	0	1	1	0
category_is_Board_ratio	aplication list	0	1	1	0	0
category_is_Board_sum	aplication list	0	0	1	0	0
category_is_Books & Reference_ratio	aplication list	1	0	0	1	0
category_is_Books & Reference_sum	aplication list	0	0	0	1	1
category_is_Business_ratio	aplication list	1	1	1	0	1
category_is_Business_sum	aplication list	0	0	0	0	0
category_is_Card_ratio	aplication list	0	1	1	0	1
category_is_Card_sum	aplication list	1	1	0	0	0
category_is_Casino_ratio	aplication list	1	0	0	1	0
category_is_Casino_sum	aplication list	0	0	0	0	0
category_is_Casual_ratio	aplication list	1	0	0	0	0
category_is_Casual_sum	aplication list	1	0	0	0	1
category_is_Comics_ratio	aplication list	1	0	1	1	1
category_is_Comics_sum	aplication list	1	1	0	1	1
category_is_Communication_ratio	aplication list	1	0	1	0	0
category_is_Communication_sum	aplication list	1	1	1	0	0
category_is_Dating_ratio	aplication list	0	1	0	0	1
category_is_Dating_sum	aplication list	1	0	0	1	0
category_is_Education_ratio	aplication list	1	0	1	0	1
category_is_Education_sum	aplication list	0	0	0	1	0
category_is_Educational_ratio	aplication list	1	0	0	1	0
category_is_Educational_sum	aplication list	1	0	0	0	0
category_is_Entertainment_ratio	aplication list	1	0	0	1	1
category_is_Entertainment_sum	aplication list	1	0	0	1	0
category_is_Events_ratio	aplication list	1	0	0	1	0
category_is_Events_sum	aplication list	0	1	0	0	0
category_is_Finance_ratio	aplication list	1	1	0	0	1
category_is_Finance_sum	aplication list	0	0	1	1	0
category_is_Food & Drink_ratio	aplication list	0	1	1	1	1
category_is_Food & Drink_sum	aplication list	1	1	1	1	0
category_is_Health & Fitness_ratio	aplication list	1	1	0	1	0
category_is_Health & Fitness_sum	aplication list	1	1	1	0	0
category_is_House & Home_ratio	aplication list	0	0	1	1	0
category_is_House & Home_sum	aplication list	0	1	0	0	0
category_is_Libraries & Demo_ratio	aplication list	1	1	1	1	1
category_is_Libraries & Demo_sum	aplication list	1	0	1	0	1
category_is_Lifestyle_ratio	aplication list	0	1	0	1	1
category_is_Lifestyle_sum	aplication list	0	0	1	1	0
category_is_Maps & Navigation_ratio	aplication list	0	1	1	1	1
category_is_Maps & Navigation_sum	aplication list	0	1	0	0	1
category_is_Medical_ratio	aplication list	1	1	0	1	1
category_is_Medical_sum	aplication list	1	0	0	1	0
category_is_Music & Audio_ratio	aplication list	1	0	1	1	1
category_is_Music & Audio_sum	aplication list	0	0	0	1	1
category_is_Music_ratio	aplication list	1	0	1	1	1
category_is_Music_sum	aplication list	1	0	1	1	0
category_is_News & Magazines_ratio	aplication list	1	1	0	1	1
category_is_News & Magazines_sum	aplication list	0	1	1	1	1
category_is_Other_ratio	aplication list	1	0	1	1	0
category_is_Other_sum	aplication list	0	1	1	1	1
category_is_Parenting_ratio	aplication list	1	1	0	0	1
category_is_Parenting_sum	aplication list	0	1	0	1	0
category_is_Personalization_ratio	aplication list	0	1	0	0	1
category_is_Personalization_sum	aplication list	0	1	0	1	1
category_is_Photography_ratio	aplication list	1	0	1	1	0
category_is_Photography_sum	aplication list	0	0	1	0	1
category_is_Productivity_ratio	aplication list	1	1	1	0	0
category_is_Productivity_sum	aplication list	1	1	0	0	1
category_is_Puzzle_ratio	aplication list	1	0	0	1	1
category_is_Puzzle_sum	aplication list	0	0	1	0	0
category_is_Racing_ratio	aplication list	1	0	0	1	0
category_is_Racing_sum	aplication list	0	1	0	1	0
category_is_Role Playing_ratio	aplication list	0	0	0	0	0
category_is_Role Playing_sum	aplication list	1	0	0	1	1
category_is_Shopping_ratio	aplication list	0	1	0	0	1

Table 11 continued from previous page

feature	data category	E (MI)	A (MI)	O (MI)	S (MI)	C (MI)
category_is_Shopping_sum	aplication list	1	0	0	1	1
category_is_Simulation_ratio	aplication list	0	1	0	0	1
category_is_Simulation_sum	aplication list	1	1	0	1	0
category_is_Social_ratio	apllication list	1	1	0	1	1
category_is_Social_sum	application list	0	1	0	0	1
category_is_Sports_ratio	application list	0	0	0	0	1
category_is_Sports_sum	application list	0	0	0	1	1
category_is_Strategy_ratio	application list	0	0	1	0	0
category_is_Strategy_sum	application list	0	1	0	1	1
category_is_Tools_ratio	application list	0	1	0	1	0
category_is_Tools_sum	application list	1	1	1	0	1
category_is_Travel & Local_ratio	application list	1	1	1	1	0
category_is_Travel & Local_sum	application list	1	0	0	0	1
category_is_Trivia_ratio	application list	1	0	0	0	1
category_is_Trivia_sum	application list	1	0	1	1	1
category_is_Video Players &Editors_ratio	application list	1	1	1	1	0
category_is_Video Players & Editors_sum	application list	0	1	1	1	0
category_is_Weather_ratio	application list	1	0	0	0	1
category_is_Weather_sum	application list	0	0	0	0	1
category_is_Word_ratio	application list	0	0	0	1	1
category_is_Word_sum	application list	1	0	0	0	1
color_mode	phone settings	1	0	1	0	0
contact_metadata_sync_enabled	phone settings	0	1	1	0	0
contact_one_month	contact list	1	1	1	1	1
contact_one_week	contact list	0	1	0	0	0
contact_six_months	contact list	0	0	0	1	1
contact_three_months	contact list	1	0	0	0	0
contact_two_weeks	contact list	1	0	0	1	1
contacts	contact list	1	1	0	1	0
contacts_address	contact list	1	0	1	0	1
contacts_family	contact list	1	0	1	1	0
contacts_fixed	contact list	1	0	0	1	1
contacts_fixed_or_mobile	contact list	1	0	1	1	0
contacts_foreign	contact list	1	0	0	0	0
contacts_ice	contact list	0	0	0	1	1
contacts_mobile	contact list	1	1	1	1	1
contacts_pager	contact list	1	1	1	1	1
contacts_photo	contact list	0	1	1	1	1
contacts_premium_rate	contact list	0	0	0	1	1
contacts_shared_cost	contact list	0	1	1	0	1
contacts_short	contact list	1	0	0	1	0
contacts_toll_free	contact list	1	0	0	1	1
contacts_type_home	contact list	1	0	1	0	0
contacts_type_mobile	contact list	1	0	1	1	0
contacts_type_unidentified	contact list	1	1	0	1	0
contacts_type_work	contact list	1	1	1	1	0
contacts_uan	contact list	1	0	1	1	0
contacts_unknown	contact list	0	1	1	1	1
contacts_ussd	contact list	0	1	0	1	1
contacts_voicemail	contact list	1	0	1	0	0
contacts_voip	contact list	1	1	1	1	1
contacts_with_duplicated_numbers	contact list	1	0	0	0	1
count_calls	calls history	1	1	0	1	0
data_roaming	phone settings	1	0	1	0	1
density_dpi	phone settings	1	1	1	0	0
development_settings_enabled	phone settings	0	0	1	1	1
dtmf_tone_type_when_dialing	phone settings	0	0	0	0	1
dtmf_tone_when_dialing	phone settings	1	1	1	0	0
duration_calls_max	calls history	0	1	0	1	1
duration_calls_mean	calls history	0	0	1	0	1
duration_calls_median	calls history	0	1	1	1	1
duration_calls_min	calls history	1	0	0	1	1
duration_calls_std	calls history	1	0	1	1	1
duration_calls_sum	calls history	1	0	1	0	0
emails	contact list	1	0	0	0	1
end_button_behavior	phone settings	1	1	1	0	0
external_emulated	phone settings	0	1	0	0	1
font_scale	phone settings	0	1	0	1	1
free_size	phone settings	1	1	0	1	1
free_size_sd	phone settings	0	1	0	0	0
games	application list	1	1	0	1	0
games_ratio	application list	1	0	1	1	0
games_sum	application list	0	1	0	1	1
haptic_feedback_enabled	phone settings	1	1	1	0	0
hard_keyboard_hidden	phone settings	1	1	0	1	0
is_keyguard_secure	phone settings	0	1	1	1	0
keyboard	phone settings	1	1	1	1	1
late_night_ratio	application list	0	1	0	0	0
lock_pattern_enabled	phone settings	0	1	0	1	0
mean_day	application list	0	1	1	0	0
mobile_data_enabled	phone settings	1	1	0	1	0
mode_ringer_streams_affected	phone settings	1	0	1	1	0
mute_streams_affected	phone settings	0	0	1	1	1
navigation	phone settings	0	0	1	0	1

feature	data category	E (MI)	A (MI)	O (MI)	S (MI)	C (MI)
navigation_hidden	phone settings	1	0	1	0	0
nfc_is_enabled	phone settings	0	0	0	0	1
nfc_is_not_available	phone settings	0	0	1	0	0
nfc_is_not_enabled	phone settings	1	0	0	1	1
number_type_is_fixed_or_mobile	calls history	1	1	1	1	1
number_type_is_fixed_or_mobile_ratio	calls history	1	0	1	1	0
number_type_is_fixed_ratio	calls history	1	1	1	1	0
number_type_is_fixed-line	calls history	0	1	0	1	1
number_type_is_FOREIGN	calls history	1	1	0	1	1
number_type_is_FOREIGN_ratio	calls history	0	1	0	1	0
number_type_is_MOBILE	calls history	1	1	0	1	1
number_type_is_MOBILE_ratio	calls history	1	0	1	1	0
number_type_is_PAGER	calls history	1	1	0	1	0
number_type_is_PAGER_ratio	calls history	1	1	1	0	1
number_type_is_PREMIUM_RATE	calls history	0	0	1	1	0
number_type_is_PREMIUM_RATE_ratio	calls history	0	0	0	1	1
number_type_is_SHARED_COST	calls history	1	1	0	0	0
number_type_is_SHARED_COST_ratio	calls history	1	0	0	1	1
number_type_is_SHORT	calls history	1	1	0	0	0
number_type_is_SHORT_ratio	calls history	1	1	1	0	0
number_type_is_TOLL_FREE	calls history	0	1	1	0	0
number_type_is_TOLL_FREE_ratio	calls history	0	1	1	1	1
number_type_is_UAN	calls history	0	1	1	1	1
number_type_is_UAN_ratio	calls history	0	0	0	1	0
number_type_is_UNKNOWN	calls history	1	1	1	0	1
number_type_is_UNKNOWN_ratio	calls history	1	1	1	0	1
number_type_is_VOIP	calls history	0	1	0	0	1
number_type_is_VOIP_ratio	calls history	1	1	1	1	1
ratio_apps_per_day	application list	0	0	0	1	1
ratio_apps_per_week	application list	0	0	0	1	1
ratio_days_per_app	application list	0	0	0	1	1
ratio_weeks_per_app	application list	0	0	0	1	1
ringer_mode_is_normal	phone settings	0	0	0	0	0
ringer_mode_is_silent	phone settings	0	1	1	0	1
ringer_mode_is_vibrate	phone settings	0	0	1	1	1
rotation_enabled	phone settings	0	1	0	1	0
screen_brightness	phone settings	1	0	1	0	1
screen_brightness_mode	phone settings	0	0	1	1	0
screen_layout	phone settings	1	1	1	0	0
screen_off_timeout	phone settings	1	0	0	0	1
sound_effects_enabled	phone settings	0	1	1	0	1
stay_on_while_plugged_in	phone settings	1	0	0	1	0
system_apps	aplication list	0	0	1	0	1
text_auto_caps	phone settings	1	1	1	0	1
text_auto_punctuate	phone settings	1	1	0	0	1
text_show_password	phone settings	0	0	0	0	1
time_12_24	phone settings	1	0	1	1	1
total_bytes	phone settings	0	1	0	0	0
total_bytes_sd	phone settings	0	1	1	0	0
total_size	phone settings	0	1	0	1	1
total_size_sd	phone settings	0	1	1	0	0
transition_animation_scale	phone settings	1	0	0	1	1
tts_default_pitch	phone settings	0	1	1	0	1
tts_default_rate	phone settings	1	1	1	0	1
type_is_BLOCKED	calls history	1	1	1	1	1
type_is_BLOCKED_ratio	calls history	1	0	1	1	0
type_is_INCOMING	calls history	1	0	1	1	0
type_is_INCOMING_ratio	calls history	0	1	1	0	0
type_is_MISSED	calls history	0	0	0	0	0
type_is_MISSED_ratio	calls history	1	1	0	0	1
type_is_OUTGOING	calls history	1	1	1	0	0
type_is_OUTGOING_ratio	calls history	1	0	0	1	0
type_is_REJECTED	calls history	0	0	0	1	1
type_is_REJECTED_ratio	calls history	1	0	1	0	0
type_is_UNKNOWN	calls history	1	0	0	0	0
type_is_UNKNOWN_ratio	calls history	1	1	0	0	0
ui_mode	phone settings	1	0	0	1	0
user_apps	application list	0	0	0	1	0
user_rotation	phone settings	0	0	0	0	0
version_sdk	phone settings	0	1	1	1	0
vibrate_when_ringing	phone settings	1	1	1	1	1
weekend_ratio	application list	1	1	1	1	1
wifi_on	phone settings	0	0	1	1	1
wifi_sleep_policy	phone settings	1	0	1	0	0
wifi_watchdog_on	phone settings	1	1	1	1	1
window_animation_scale	phone settings	1	1	1	0	0
work_period_ratio	application list	1	1	0	0	0