1 2 9 0

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Mafalda Viana Bastos

Personality Detection From the Use of Smartphone

Thesis submitted to the
University of Coimbra for the degree of
Master in Biomedical Engineering

Supervisors:
Prof. Dr. Luís Macedo
Prof. Dr. Amílcar Cardoso

Coimbra, 2019

Resumo

Numa era cada vez mais tecnológica, a utilização e comercialização de *smartphones* emergiu exponencialmente na última década. Hoje em dia, existem mais de 2 mil milhões de pessoas que utilizam o seu *smartphone* diariamente, tornando-se este um dos objetos mais pessoais que temos connosco a qualquer momento. Tal permite que todas as ações e comportamentos possam ser capturados de uma forma não intrusiva e sem precedentes o que, apesar de parecer uma desvantagem da vida moderna, pode trazer-nos muitos benefícios no futuro.

O objetivo deste estudo é assim perceber se é possível detetar a personalidade dos utilizadores a partir do seu comportamento. Isto porque a deteção automática da personalidade através do telemóvel pode ter muitas aplicações, nomeadamente publicidade, gestão de recursos humanos, deteção precoce de depressão e outras doenças do foro psicológico, melhoria de Assistentes Pessoais Inteligentes, entre outros.

Para alcançar este objetivo, o problema é abordado através de técnicas de *machine learning*, nomeadamente algoritmos de classificação. Em primeira instância é utilizada classificação binária e, em segunda instância, com várias classes. Em ambos os casos foram desenvolvidos 5 classificadores, em que cada um visa prever um dos 5 principais traços de personalidade: Agradabilidade, Conscienciosidade, Estabilidade Emocional, Extroversão e Abertura à Experiência. Estes classificadores pretendem prever cada traço de personalidade a partir da utilização de aplicações móveis.

Com uma *accuracy* 22 % melhor que aleatório, os classificadores binários mostraram ser duas vezes melhor que os classificadores com várias classes. A Estabilidade Emocional e Extroversão são os traços mais fáceis de prever, tendo ambos os modelos uma *accuracy* de 67,5 %. Estes modelos podem beneficiar bastante a capacidade de adaptação dos Assistentes Pessoais Inteligentes aos seus utilizadores e, se desenvolvidos, podem-se tornar úteis em muitas outras áreas.

Palavras-Chave: Aprendizagem Máquina, Classificadores, Deteção da Personalidade, *Big Five*, Telemóveis, Uso de Aplicações Móveis, Comportamento

Abstract

In an increasingly technological era, the use of smartphones has emerged exponentially in the last decade. Nowadays, over 2 billion people use their smartphones on a daily basis, making them one of the most personal objects we have with us at all times. This allows these devices to capture our actions and behaviors in an unobtrusive and unprecedented way. Despite seeming like a disadvantage of modern life, this knowledge can bring us many benefits in the future.

The goal of this study is to understand if it is possible to identify smartphone users' personality from their behavior. This automatic detection of personality can have numerous applications, such as advertising, human resources management, early detection of depression and other psychological illnesses, improvement of Intelligent Personal Assistants (IPAs), among others.

To achieve this goal, a machine learning approach is used based on classification algorithms. This approach is divided into two parts: binary classification and multi-label classification. In both cases, five classifiers were created in order to predict each of the Big Five personality traits: Agreeableness, Conscientiousness, Emotional Stability, Extroversion, and Openness. These classifiers aim at predicting each trait from mobile applications usage through smartphones.

The binary classifiers proved to be two times better than the multi-label ones, with an accuracy 22% better than random. Emotional Stability and Extroversion are easier to predict than other traits, with both models reporting an accuracy of 67,5%. These models can greatly benefit IPAs' capability of adapting to their users and, if further developed, can be useful in many other fields.

Keywords: Machine Learning, Classifiers, Personality Detection, Big Five, Smartphones, App Usage, Behavior

Acknowledgments

First of all, I would like to thank my supervisor, Luís Macedo, for all the support he provided throughout this project. He handed me the liberty and responsibility to investigate what I truly wanted, offering me a unique opportunity for academic growth.

Board of European Students of Technology was one of the most surprisingly positive experiences I had the pleasure to be a part of during the last five years. Not only did it contribute to most of the soft skills I acquired during my time at this University but also allowed me to meet amazing and broad-minded people all over Europe. I'm thankful for all the incredible challenges it put me through and the personal growth they provided me.

To all my friends, with whom I shared the last five years, thank you for all the amazing memories. From the late hour projects and stressful exam seasons to the amazing travels and unique academic traditions, I am grateful that I was able to share them all with you. I hope we'll continue to be in each other's lives and that, wherever life takes us, we will always celebrate each other's achievements. Just remember that if the rain starts to fall "I'll be there for you like I've been there before. I'll be there for you because you've been there for me too."

Finally, to my family, especially my mother, to whom I will always be grateful for being my biggest support. I would not be who I am today without your firm hand and unconditional love. You have always been my safe harbor, allowing me the freedom to be who I am without ever letting my feet drift too far off the ground. Thank you for everything.

"*If we knew what we were doing, it would not be called research, would it?*"

Albert Einstein

Contents

List of Tables

List of Figures

Abbreviations

AI Artificial Intelligence.

BFI Big Five Personality Inventory.
BFI-10 Big Five Inventory-10.
BFSI Big Five Structure Inventory.
BOOST Gradient Boosting.

CNN Convolutional Neural Network.

EDA Exploratory Data Analysis.

GPS Global Positioning System.

IPAs Intelligent Personal Assistants.

KNN K-Nearest Neighbors.

LDA Linear Discriminant Analysis.
LR Logistic Regression.

NB Gaussian Naive Bayes.
NEO PI-R The Revised NEO Personality Inventory.
NEO PI The 60-item NEO Personality Inventory.

PCA Principal Component Analysis.

RF Random Forest.

SVM Support Vector Machine.

TIPI Ten-Item Personality Inventory.

1

Introduction

1.1 Motivation

Dealing with humans is one of the biggest challenges in computing, whether they are users, subjects in the data to be analyzed or digital material producers and consumers. Personality, as a construct capable of capturing the most evident characteristics of an individual, might provide the key to better bridge the gap between people and machines [3]. This will allow us to further improve Intelligent Personal Assistants (IPAs), in which Artificial Intelligence techniques are drawn upon to develop systems capable of providing people with more tailored assistance. Examples of such a system would be Apple Siri, Amazon Alexa, Google Assistant or Microsoft Cortana [4].

IPAs are software agents that can automate and ease many of the daily tasks of their users [5]. By gathering knowledge and awareness about the usual behavior of their users while interacting with them and/or collecting data, IPAs are supposed to adapt themselves to their user's needs and actions, in order to improve the given assistance [6]. In this way, IPAs can monitor the behavior of an individual and produce models of what the individuals know, how they feel, what are their motives, intentions, and desires and later predict for specific contexts the mental states of their users and act accordingly.

Knowing its user's personality would greatly benefit IPAs from two different perspectives. First, individuals with different personality display different kinds of behavioral patterns. With this information, IPAs would be able to make informed guesses and adapt to their user at an early stage, providing much more valuable support. Second, IPAs could benefit from synthetic personalities, compatible with its user's. This would increase their acceptance, especially among people that are not very familiarized with technology [7].

Besides the additional value that predicting someone's personality would bring to IPAs, there are many more applications for this area of research. From a sales point of view, by relating everyday behavior with personality, it might be possible to target advertising campaigns to the right potential customers [8] and predict the success of a new product [3].

Furthermore, automatic personality detection and its relation to everyday behavior can also play a major role in the medical field. Diseases like paranoia and schizophrenia typically interfere with personality [9], with patients reporting a very specific set of personality traits. Automatic detection of these cases allows the identification of people at risk which, with additional medical exams, can lead to an easier and earlier diagnose. Moreover, promising research based on smartphone-related behaviors has focused on predicting stress [10] and depression [11, 12, 13], studying bipolar disorder [14, 15] and smartphone addiction [16].

1.2 Goals

The main goal of this study is to determine whether it is possible to predict someone's personality traits based on their activity on their smartphone, specifically their mobile app usage. The data set publicly available at [17] is used to achieve this goal.

In the first place, exploratory data analysis must be implemented as well as feature selection techniques. The aim is to understand which features are more relevant for each personality trait and explore the meaning of the obtained results.

Based on the previous step, supervised learning algorithms are implemented in order to try to predict smartphone users' personality traits. The problem is approached as a classification problem with the intention of determining if people tend to manifest the most positive or negative pole of each personality trait. With this in mind, two experiments are performed: Binary Classification and Multi-class Classification. On both experiments, the most important features for each personality trait are taken into consideration and five classifiers are built, one for each of these traits (a full description of such experiments is presented in Chapter 4).

Afterwards, the models are evaluated in order to assess their success in predicting someone's personality traits based on their behavior towards mobile applications. In addition, the two experiments are analyzed and compared.

1.3 Scientific Contributions

Although there is a growing interest in the scientific community to take advantage of the information smartphones can provide, few studies focus on predicting personality traits from smartphone usage.

This thesis contributes to the field of personality detection in different aspects. Firstly, a publicly available data set is used, making it possible to replicate results and allowing an unbiased comparison of different methodological approaches.

Secondly, this work leverages actual behavior data by using data collected in a real-life context. This avoids the classical approach that relies on self-reported behavior questionnaires, which are challenging and time-consuming, commonly used in the field of psychology.

Moreover, unlike previous studies that only consider the use of some mobile applications (apps), the data used in this work contains information on all the main app categories present in the different app stores, thus allowing a thorough comparison between them. The recorded information about these apps is also more comprehensive than in previous research since it tracks both the frequency and the duration of use of each app category while others only track app installation. This reflects better the daily behavior of participants since it is common for people to install apps but seldom use them.

Finally, this work confirms previous findings claiming that personality traits have an impact on smartphone usage and it is able to develop a machine learning approach, based on classification, that would allow IPAs to improve their capability of adapting to their users. This approach can also be reproduced by future studies since it is easily escalated to larger data sets. To the best of our knowledge, this is also the first work to compare a binary and multi-class classification for personality detection.

1.4 Document Structure

The remainder of this document is structured as follows: Chapter 2 provides some general insight about the basis of this work. In Chapter 3, previous research on personality detection is analyzed as well as studies relating personality and smartphone behavior. Then, Chapter 4 provides an overview of the different stages of this work and the machine learning methods used to address the main goals already established in Section 1.2. In Chapter 5, the obtained results are presented and discussed. Binary and multi-class classification are also compared. Finally, Chapter 6 concludes the thesis, providing some insight about the findings of this research as well as its limitations. In addition, possible directions for future improvements are provided.

2

Background Knowledge

2.1 Personality

Personality aims at capturing stable individual characteristics, typically measured in quantitative terms, that explain and predict observable behavioral differences [18]. Personality has puzzled scientists and philosophers over the centuries, which ultimately lead to the appearance of Personality Psychology. Some define this field as: "Personality is that branch of psychology which is concerned with providing a systematic account of the ways in which individuals differ from one another" [19]. The key-assumption of Personality Psychology is that stable individual characteristics result in stable behavioral patterns that people tend to display, at least to a certain extent, independently of the situation [3]. Thus, personality can be seen as one of the several factors that influence our behavior at all times.

Different personality models are accepted by the scientific community but the ones that most effectively predict measurable aspects in the life of people are those based on traits [3]. Trait models build upon human judgments about semantic similarities and relationships between adjectives that people use to describe themselves and the others [20]. These models are also widely accepted in the computing community since they represent personality in terms of numerical values, a form particularly suitable for computer processing [7].

There are several prominent trait models in the literature, including Allport's Trait Theory [21], Cattell's Sixteen Factor Model [22], Eysenck's Giant Three [23], and the Myers–Briggs Type Indicator [24]. Notwithstanding, the Big Five Model [25], which is a five-factor model, is the most dominant trait model of our time in the field of Personality Psychology [26]. In addition, it is also the most commonly adopted model in computing oriented research, which reflects how widely accepted the Big Five model is by the scientific community [3].

The Big Five Model does not aim at sorting anybody into a "type" but rather inform on where someone falls on a continuum of personality traits. Each of the five traits is evaluated independently and points to a range between two extremes. In no particular order, the Big Five personality traits are Neuroticism, Extroversion, Openness, Conscientiousness, and Agreeableness [27]. Figure 2.1 summarizes the characteristics of these five personality traits.

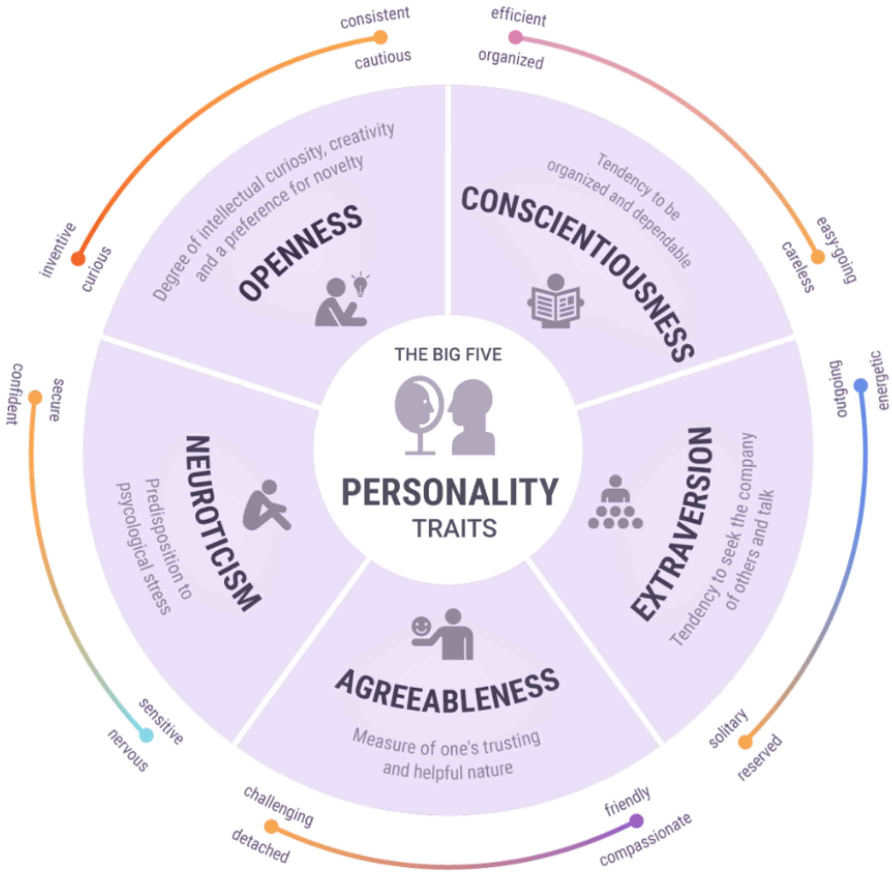


Figure 2.1: The Big Five personality traits (adapted from [1]).

Neuroticism is generally related to a predisposition to psychological stress [25]. It is also commonly referred to as Emotional Stability since the two terms are simply labels for the positive and negative poles of the same construct [28]. Neuroticism is related to being anxious, depressed, angry, embarrassed, emotional, worried, and insecure [29]. On the other end of the spectrum, people who tend to be emotionally stable are usually calm and confident.

Extroversion is considered a tendency to seek the company of others and to be communicative [25]. People who score high in Extroversion tend to be social, assertive, outgoing, and talkative. They usually place a high value in close interpersonal relationships [30]. In social settings, it is common for extroverts to appear more dominant than introverts, who tend to be less talkative and more reserved [25].

Openness is the degree of intellectual curiosity, creativity, and a preference for novelty [25]. It is frequently associated with being imaginative, cultured, curious, original, broad-minded, intelligent, and artistically sensitive [29]. On the other hand, people who score low on this trait tend to be more cautious and realist.

Conscientiousness assesses the tendency to be organized and dependable [25]. People who score high on this trait are linked with being careful, thorough, responsible, organized, and planful [29]. They are self-disciplined and success-driven [31] thus being associated with a high performance in study and jobs [32]. Contrarily, people who score lower on this trait are more flexible and easy-going which, at its extreme, might come across as being careless [25].

Lastly, Agreeableness is a measure of one's trusting and helpful nature [25]. This trait is linked with being courteous, friendly, trusting, good-natured, cooperative, forgiving, soft-hearted, and tolerant [29]. People who score low on this trait often come across as detached and untrustworthy [25].

In addition to these five factors, much research on the Big Five has focused on a two-level hierarchy, with the five factors at the top containing narrower traits called "facets" at a second level [33]. Facets are thus considered a specific and unique aspect of a broader personality trait [34] and each of the Big Five traits contains six facets [35], which can be seen in Table A.1. These are helpful to further understand what information each personality trait comprehends as well as to assess specific characteristics within each trait.

It is interesting to note that there is a strong genetic component that influences people's personality and that studies have shown that personality is partially inherited [36, 37]. Everyone's unique cluster of genetic traits predisposes them to a particular personality mainly because all personality traits have biological underpinnings [38]. People who score high in Agreeableness, for example, tend to have higher natural oxytocin levels, a hormone that plays an important role in social bonding, increasing feelings of trust, and that acts as a natural anti-depressant [39, 40]. Neuroticism, on the other hand, has been linked to a hypersensitivity of the amygdala, the portion of the brain responsible for noticing threats [41]. This means that people who are more sensitive to negative cues of the environment, tend to score lower on Emotional Stability.

The Big Five personality traits are usually assessed by performing questionnaire-based personality tests. A numerous amount of different personality tests can be found in the literature. Some examples are listed below:

- Big Five Personality Inventory (BFI) [27].
- Big Five Inventory-10 (BFI-10) [42].
- Big Five Structure Inventory (BFSI) [43].
- The 60-item NEO Personality Inventory (NEO PI) [44].
- The Revised NEO Personality Inventory (NEO PI-R) [45, 46].
- Ten-Item Personality Inventory (TIPI) [47].

Even though all these tests are questionnaire-based and focus on the Big Five personality traits, they differ in the way they assess them. Regarding the questions asked, they differ both in type and in quantity. In addition, the range of the personality trait scores obtained varies from test to test, making it difficult to compare results between them.

These tests are useful because they allow people to know their personality, which is fairly stable in adults [48] and can have an impact on an individual's decision-making process [49]. Other studies [50, 51] also show that individuals structure their habits, lifestyle and general interactions with the environment partially based on personality. Furthermore, many marketing managers also believe that consumers buy and use products according to their personalities [52].

Recent studies claim that Personality Psychology can benefit from the current digital age [53] and that this field has never been in better health than in the last decade [54]. This growing knowledge about the subject is important so that we can understand our behavior, emotions and motivations and its automatic detection has many practical applications in medicine, forensics, sales and advertising, human resource management [55], among many others.

2.2 Smartphones

Smartphones represent an important part of modern life, allowing us to communicate with others from nearly anywhere, to access the internet and check several different mobile applications (apps) [56]. Nowadays, they are the most personal devices people own and carry around with them all day [49].

In the last decade, advanced economies reached high rates of smartphone ownership and usage. However, recent studies reveal that emerging economies are catching up with this reality, with a smartphone ownership rate that continues to climb [57]. This worldwide growth of smartphone usage provides a new lens for investigating mobile phone usage [58]. It offers unprecedented opportunities for tracking, monitoring and documenting people's movements, behavior, environment, biosignals and much more in real time and across different contexts. Additionally, smartphones also mediate social interactions, making it possible to assume that smartphone usage could actually reflect an individual's personality [59].

All this information, related to human behavior, that smartphones are able to capture, is the focus of many studies in the social, health and behavioral sciences. However, few studies take advantage of these unobtrusive observation methods that collect objective measures of daily life behavior, out in the real world, which is supposedly the context of ultimate interest [60]. Smartphones allow us to register and observe all this information regarding an individual's behavior without interfering with it. This is a great advantage for the field of Personality Psychology since most studies in this area used to be focused on self-reported behavior, neglecting observable acts [61]. In addition, individuals may intentionally or unintentionally under-report or over-estimate some of their behaviors [62], which might influence the results obtained. Thus, this former approach based on self-reported behavior has been criticized by different authors [60, 61]. Besides, research [63] shows that to fully understand personality and real-life behavior, one must consider them outside a controlled environment, like a laboratory. In this way, smartphones can provide the necessary tools to significantly improve any research that takes into account human behavior.

The global use of smartphones has also made an impact on the mobile app industry. A mobile application, most commonly referred to as an app, is a type of application software designed to run on a mobile device, such as a smartphone or a tablet [64]. By the end of 2018, there were over two million apps for download in the leading app stores [65]. This enormous number of mobile apps allows users to customize their experience on their smartphones according to their interests, creating fertile ground to study the behavior of smartphone's users [10]. Moreover, previous studies found that there is an overall diversity in how different people use smartphones [66] and that it is possible to differentiate users through their set of used apps, their *app signature* [67]. The kind of apps people install and use could be related to their interests, demographics, and personality [68, 69, 70].

2.3 Artificial Intelligence and Machine Learning

In order to establish a relationship between personality and smartphones, Machine Learning techniques are used. Therefore, it becomes relevant to define Artificial Intelligence and Machine Learning, as well as understand their basic components.

Russell & Norvig (2010) consider Artificial Intelligence (AI), which started to be developed soon after World War II, to be one of the most modern fields in science and engineering [71]. AI can be seen as the broader concept of a system's ability to correctly interpret external data, to learn from it, and to use those learnings to achieve specific goals and tasks through flexible adaptation [72]. Machine Learning, which is seen as a subset of AI, is the science of programming computers so they can learn from data [73]. In turn, a machine learning model can be considered a mathematical function that aims at reproducing a real-world process [74]. In the past decade, this field has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome [75].

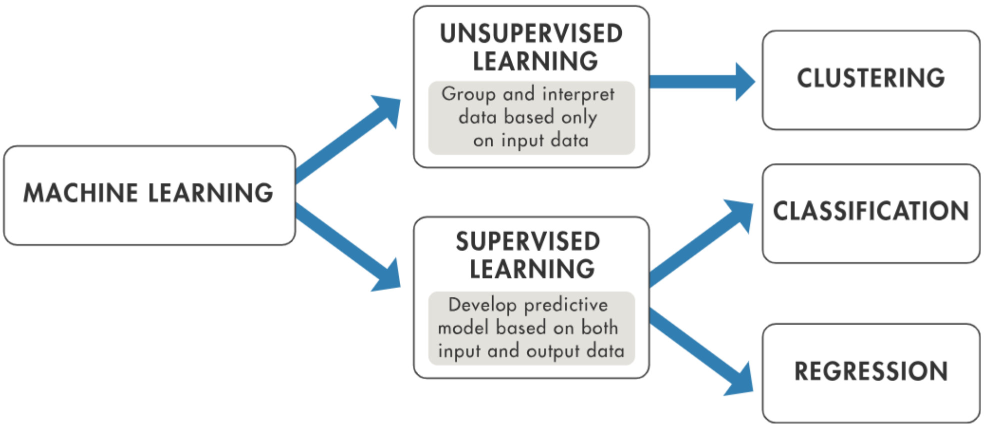


Figure 2.2: Machine Learning Techniques (adapted from [2]).

As compiled in Figure 2.2, machine learning uses two main types of techniques: supervised learning and unsupervised learning [2]. A supervised learning algorithm takes a known set of input data and known responses to the data, the output, and trains a model to generate reasonable predictions for the response of new data. More formally, considering the pair $(x, f(x))$ an example, x is the input and $f(x)$ is the output of the function applied to x. The goal of supervised learning is: given a collection of examples of f, return a function h that approximates f [76].

Supervised learning can resort to either classification or regression techniques to develop predictive models [77].

Classification techniques predict discrete responses, classifying the input data into categories. These categories are also commonly referred to as classes or targets. Thus, a classification model can be able to predict whether an individual is an introvert or an extrovert.

On the other hand, regression techniques predict continuous responses. Thus, a regression model can be able to predict an individual's score on a personality test if this value belongs to a known data range.

Unlike the previous type of algorithms, unsupervised learning does not need the output of the data it receives. It finds hidden patterns or intrinsic structures based only on the input data [77] which can be helpful, for example, when performing Exploratory Data Analysis (EDA). Clustering, which aims at detecting potentially useful clusters of input data, is an example of these types of algorithms [71].

2.3.1 Used Technologies

As aforementioned, the core of this work is based on supervised learning algorithms and classification techniques. These were implemented in Python using Scikit-learn [78].

2.3.1.1 Python

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability. Its language constructs and object-oriented approach aims at helping programmers write clear, logical code for small and large-scale projects [79].

Nowadays, Python is one of the most popular programming languages for Machine Learning problems [80]. One of the reasons for this growing popularity is its great library ecosystem. These libraries are modules, published by different sources, that include a pre-written piece of code that allows users to reach some functionality or perform different actions without having to code them from the very beginning every time [80].

2.3.1.2 Scikit-learn

Scikit-learn is one of the most well-known libraries in Python and it integrates a wide range of state-of-the-art machine learning algorithms for supervised and unsupervised problems. The goal of this package is to make machine learning accessible to non-specialists through a general-purpose high-level language [78]. Further information can be obtained from Scikit-learn's official website (`https://scikit-learn.org`).

3

State of the Art

As already discussed, personality is typically measured using quantitative tests. However, since the automatic detection of personality can have so many benefits, efforts have been made in the last two decades to further advance this field of research. Taking advantage of the Digital Age we live in, numerous studies have tried to relate personality and behavior.

This chapter begins by taking a more comprehensive look at research on personality detection in general, in Section 3.1. Afterwards, a more in-depth analysis of the relationship between personality and smartphone usage is provided in Section 3.2. In the end, Section 3.3 provides a brief analysis of the research mentioned throughout this chapter.

3.1 Personality Detection

Social Media platforms mediate a major part of modern-day social interactions, which has made them the focus of a significant amount of research that aims at analyzing personality and behavior. The platforms taken into consideration in this Section include Facebook, Youtube, Twitter, and Instagram. In the end, personality detection through text is also addressed.

Focusing on social media use, Correa et al. (2009) investigated the relationship between three traits of the Big Five Model and the use of social networking sites and instant messages [81]. Age and gender were also taken into consideration for this analysis. The results revealed that while Extroversion and Openness were positively related to increased social media use, Emotional Stability was a negative predictor.

Nowadays, it is common for people to only want to share on social media their ideal-self instead of their real-self. To study this, Back et al. (2010) asked a number of participants to rate their ideal-self. These rating were then compared to strangers' ratings of participants based solely on the participants' user profiles and an accuracy criterion. The latter included participants' self-ratings and ratings of the participant by multiple informants who knew them offline. The findings showed that strangers' ratings correlated strongly with the accuracy criterion but weakly with the ideal-self ratings, suggesting that online social networking profiles reflect fairly accurate personality traits of profile owners [63].

By showing that social networks tend to reflect the actual personality of its users, this research laid the foundation for others, encouraging them to take advantage of these platforms to study personality detection.

Yeo (2010) analyzed the relationship between of the Big Five personality traits and Youtube usage. It was concluded that the use of this application was indeed influenced by the personality of its users, with Extroversion and Openness being the ones with the most significant effects [82].

Focusing on Instagram usage, which allows photo-sharing and applying different photo filters, Ferwerda et al. (2015) tried to infer personality traits from this social network usage. The results revealed a significant correlation between picture features and personality traits, leading the authors to believe that personality impacts the way people want their pictures to look [83].

Twitter, often considered a micro-blogging service, is a social network on which users post and interact with messages known as "tweets". By taking into consideration only the number of following, followers, and listed counts on Twitter, Quercia et al. (2011) showed that it is possible to accurately predict a user's personality. It was found, among other things, that popular users (i.e. those who are followed by many) typically score high in Extroversion, Emotional Stability and Openness [84]. Qiu et al. (2012) showed that tweets contain valid linguistic cues to personality that can be used by other people to reliably detect Agreeableness and Neuroticism [85].

More recently, Liu et al. (2016) developed deep-learning based models that, when applied to tweets, shows state-of-the art performance across five personality traits and three languages (English, Spanish and Italian) [86]. On the other hand, Carducci et al. (2018) proposed a supervised learning approach to also predict personality traits from tweets [87].

Skowron et al. (2016) defend that by collecting information from two distinct social networks, one based on text (Twitter) and another based on image (Instagram), a more comprehensive profile of the user's personality can be obtained. By resorting to this approach, the findings showed a consistent decrease of the prediction errors for each personality trait [88].

Furthermore, there is a growing interest in the scientific community to take advantage of written text in general, not only tweets, for personality detection. This is probably due to the fact that language psychology shows that the choice of words is driven not only by meaning, but also by psychological factors such as emotions, relational attitudes, power status and personality traits [3].

Argamon et al. (2004) aimed at differentiating high from low Neuroticism and Extroversion in informal text. The results obtained using a Support Vector Machine algorithm suggest that the best predictor for Neuroticism are appraisal adjectives and modifiers, and that standard function words work best for Extroversion [89].

On the other hand, Mairesse et al. (2007) have used classification, regression and ranking models to try to predict the Big Five personality traits from both conversation and text. The findings show that personality can, indeed, be recognized by computers through language cues, with the best results being obtained with ranking models [90].

More recently, Majumder et al. (2017) have used a deep learning approach to assess the Big Five personality traits from stream-of-consciousness essays. Each of the Big Five traits were predicted through an individual Convolutional Neural Network (CNN) with a binary output [55].

3.2 Personality and Smartphone Usage

Different authors assess the relationship between personality and smartphone ownership and usage by collecting self-reports of behavior and relating them to personality inventories scores.

This method was used by Lane & Manner (2011), that counted with 448 people participating in their study. Respondents who owned a phone were asked to rate from 1 to 5 the importance of different smartphone functions. Later, it was concluded that extroverts were more likely to own a smartphone and reported greater importance on the texting function. Also, more agreeable participants favored the use of smartphones for calls much more than for texting [91].

In a different study, Kim et al. (2015) adopted a similar method but, in contrast with the previous, participants were asked to complete a survey indicating whether or not a particular mobile application was used. A survey on sociodemographic variables and a personality test were also performed. It was reported that the personality traits were moderately intercorrelated, with more open individuals tending to be more extroverted and less conscientious. On the other hand, agreeable participants tended to be more extroverted and emotionally stable. Turning towards the relation between behavior and personality, Extroversion and openness were linked with an increased probability of owning a phone. Furthermore, Extroversion was related to an increased use of social networking and instant messaging applications, and Conscientiousness was linked with a decreased use of finance and shopping apps [92].

Turning towards the ability of smartphones to automatically retrieve information, the following studies are able to take advantage of smartphones' features by relating the behavioral patterns they obtain to personality tests scores.

Focusing on phone logs, Montjoye et al. (2013) used a classification-based approach in order to predict personality traits. The indicators used as predictors were: basic phone usage (e.g., number of calls, number of texts), active user behaviors (e.g., number of calls initiated, time to answer a texts), location (e.g., radius of gyration, number of places from which calls have been made), regularity (e.g., temporal calling routine) and diversity (e.g., call entropy). A Support Vector Machine (SVM) classifier was used to predict whether the users scored low, medium or high in each of the personality traits [93].

Chittaranjan et al. (2011) resorted to software to automatically collect anonymized logs of calls, text messages, Bluetooth scans and application usage of 83 participants over a period of 8 months. The study concluded that the features obtained from smartphones could be an indicator of the Big Five personality traits. Moreover, it was found that Conscientiousness was linked with a higher use of emails, and text message apps but a lower use of audio, video or music apps. Interestingly, participants who lacked in Emotional Stability also reported a higher use in email apps. On the other hand, introverts were less likely to use internet apps at all and participants with a lower score on both openness and Agreeableness reported higher usage of text message apps. Additionally, five supervised learning classifiers were built in order to predict each of the Big Five personality traits from smartphone usage. SVM and C4.5 algorithms were used and both performed above the level of chance. Nevertheless, it must be taken into consideration that most of the participants did not own a smartphone before the study. As a result, most of the features of the phone were discovered during the study which might have influenced the results [94]. The same authors later published another study [95] further analyzing the same problem, this time with a bigger data set. Small changes were applied to the machine learning model approach and gender was taken into consideration for this study.

Stachl et al. (2017) investigated to what degree personality traits, fluid intelligence, and demographic variables were able to predict the frequency and duration of mobile app usage on smartphones. The results suggest that Extroversion, Conscientiousness, and Agreeableness are particularly predictive of specific behavioral categories of app usage. Extroverted people were linked with a higher number of calls and an intensive use of photography apps. Agreeableness could be related to a higher use of transportation apps, whereas high Conscientiousness was associated with a lower usage of game-related apps. On the other hand, Openness and Emotional Stability were not associated with any particular behavior towards apps. This research [61] focused on both factor and facet level and concluded that personality scores on factor level might work best for predicting categorical app usage .

Instead of tracking app usage, it is also possible to focus on the apps installed. Xu et al. (2016) confirmed that personality traits have a significant impact on the adoption of different types of mobile apps. The findings support that Extroversion is negatively associated with mobile gaming apps. Neuroticism is positively associated with the adoption of mobile photography apps and personalization apps, whereas Agreeableness is negatively linked with the last. In addition, it was found that Conscientiousness is negatively associated with the adoption of music and video, photography and personalization apps. Furthermore, a machine-learning model was developed in order to predict a user's personality based on his installed apps. A total of ten models were built using a Random Forest Algorithm. The authors defend that users with higher and lower personality traits, alternatively to users with medium scores, are better suited for the model since they behave differently from the majority [49]. Therefore, the predictive models focused on accurately classifying people in the 'High' and 'Low' groups of each of the five personality traits.

Montag et al. (2015) focused on recording only WhatsApp behavior of 2418 users over a 4 week period. The authors concluded that women and younger ages used the application for longer periods of time. Additionally, Extroversion was positively associated with daily WhatsApp usage, while Conscientiousness showed an inverse correlation with it [56].

Finally, more recently, Viana et al. (2018) used information collected from smartphone's location, calls, battery usage and charging, networking context like Bluetooth devices and Wi-Fi access points in proximity. It was found that individuals high in Conscientiousness move in repetitive and predictable ways, and commute back home at the same time every day. Someone that scores high in Extroversion has, on average, a higher number of social interactions with others. Openness, on the other hand, is connected to seeking new experiences to try and, as a result, travel further from home and visit, on average, a higher number of location [96].

3.3 Final Remarks

A significant amount of research has been done over the last years with the goal of relating personality and text as well as the use of the internet and social media. Thus, it should be noted that only a small part of this work was explored in Section 3.1. Notwithstanding, with the appearance of smartphones, there are now devices that allow the unobtrusive measure of real behavioral patterns continuously over time. This has shifted the focus, at least in part, from the web context to smartphones, with several studies beeing carried out in the recent past relating smartphone usage to personality.

The aforementioned studies focus both on self-reported behavior and behavioral patterns automatically collected. A greater emphasis is given to the latter for two main reasons. Firstly, personality detection through behavioral patterns automatically collected from smartphones is the focus of this work. Thus it becomes important to analyze similar research to be able to compare methodologies and results. Secondly, and as already discussed in Chapter 2, self-reported behavior obtained through questionnaires, which doesn't take advantage of the above-mentioned benefits of smartphones, has several disadvantages. Nevertheless, these studies provide interesting findings, useful for future result comparison.

Regarding the studies that focus on the behavioral patterns smartphones obtain and personality traits, and to summarize, the above examples highlight that the two can be related, encouraging additional research in the field. However, most of the above-mentioned studies focus on predicting behavior while the present work aims at predicting personality.

In addition, the studies that do focus on predicting personality from smartphone usage are still recent and sparse, revealing opportunities for further improvement. The conclusiveness of their results is also limited and shouldn't be generalized since they show significant methodological differences. Different personality tests are used and, due to the extensive information that can now be retrieved from smartphones, the features taken into consideration vary a lot from study to study, making it hard to compare conclusions.

Notwithstanding, by taking an overall look at the work mentioned in this chapter, it is interesting to note that every study, both in Section 3.1 and Section 3.2, resorts to the Big Five personality traits. Once again, this shows how widely accepted the Big Five model is in the field of Personality Psychology.

4

Methodology

This chapter focuses on the steps taken in order to achieve the final machine learning models. Two different approaches, both based on classification, are taken: firstly, for each personality trait, two classes are created (high and low); secondly, for each personality trait, three classes are created (high, medium and low).

Both experiments focus on classification due to two main reasons:

1. There are numerous personality tests, with each of them assessing personality in a different way, ultimately leading to a wide variety of metrics and score ranges. Classification thus allows higher scalability than regression. Not only do these models adapt much easier to all types of scores ranges obtained from personality tests but also allow a better comparison between results obtained from different personality tests.

2. It is possible to know individuals' characteristics without knowing the exact score they obtain for each personality trait. By splitting participants into classes according to their scores, it is possible to know what end of the spectrum of a given personality trait they tend to represent. This information alone makes it possible to differentiate and characterize people thus enabling IPAs to adapt to their users.

Therefore, classification is considered the best approach since it fulfills the purpose of this study while allowing scalability of the models and a better comparison of the results obtained through different methodologies.

The 1^{st} Experiment performed to address this problem focuses on binary classification, which is a simple but widely used approach in this field. As described in Figure 4.1, two classes were defined using the median value of the personality traits' scores of all participants. This was done in order to be able to differentiate, for example, the more introvert participants from the more extroverted ones.

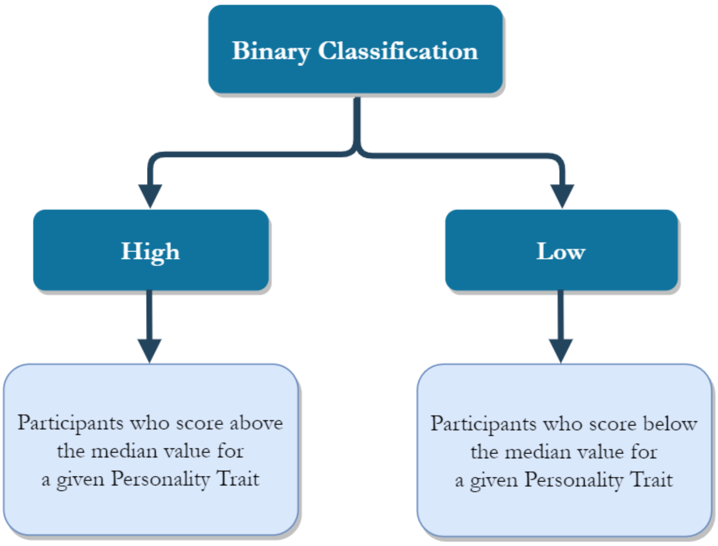


Figure 4.1: Class distribution for binary classification - 1st Experiment.

On the other hand, some authors [97] argue that the dichotomization of quantitative measures negatively impacts the analysis. In addition, personality traits have a distribution close to normal, which means that a large number of people may not be considered introvert or extrovert because they fall in the middle of the spectrum of trait scores. Thus, in the 2nd Experiment, multi-class classification is used. By splitting the participants into three classes, people in the high and low classes can show a more representative behavior of Extroversion and Introversion respectively [49]. Participants in the medium group are more prone to show average behavior that does not represent either the positive or negative pole of a personality trait. In this case, the 33rd and 66th percentile of each personality trait score were used in order to split all instances into 3 classes, as described in Figure 4.2.

In both approaches, five different classification models are built, one for each personality trait. The ultimate goal is for these models to be able to automatically predict the category to which a person belongs to, taking as input only information on smartphone usage. Ultimately, the success of the classifiers is measured according to their capability of performing such task.

To do that, the standard steps of the machine learning workflow are followed and both binary and multi-class classification are compared in Chapter 5. Data Cleaning and Formatting, and Exploratory Data Analysis (EDA) are common to both approaches while Feature Engineering and Selection, and all the steps involved in the development of the models are done separately according to each approach.

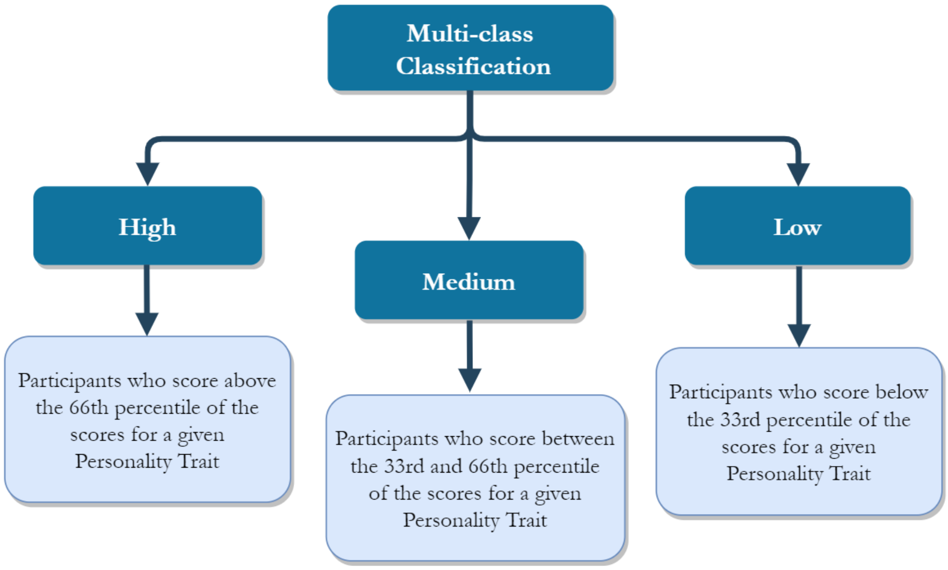


Figure 4.2: Class distribution for multi-class classification - 2nd Experiment.

4.1 Data Set

Every machine learning problem starts with a data set. This is a collection of data where, typically, every column represents a particular feature, and each row corresponds to a given instance of the data set in question. In this case, and due to the way that instances are split into two or three classes, the data set is considered to be balanced (i.e. the data set contains an equal or almost equal number of instances from each class) in both experiments.

The first step should be to analyze the data set, understand its structure and look for anomalies that require data cleaning and formatting. In this case, the data set, which is publicly available at: `http://bit.ly/stachl-data-set`, was already properly formatted which made this step easier to accomplish.

The data set used in this study [17] contains 137 instances and 94 features, which hold information on each participant. This information includes personality and fluid intelligence scores, demographic variables and user behavior recorded through an Android logging app. The last includes information on the usage frequency of each app category and the average duration of each use for all app categories. Table B.1 compiles all these categories and the corresponding mobile application they include.

The personality scores, which include information on both factor and facet level, were measured with the German version of the BFSI [43]. However, since the study of personality facets doesn't align with the goals of this study, facet scores are discarded before beginning any analysis.

4.2 Exploratory Data Analysis

Before moving to more in-depth steps of machine learning, it is important to have an overview of the data available. Thus, with the help of summary statistics and graphical representations, EDA techniques may gather patterns and characteristics, and spot anomalies in the data [98].

In order to better understand the population of this study, fluid intelligence scores and demographic information are taken into consideration when performing EDA. Nevertheless, with exception to the gender, these are not considered for the next steps described in this chapter. The goal of this work is to study the relation between personality and behavior, and not demographic variables. Hence, the data used in all the following steps includes:

- **55 independent variables:** gender, usage frequency of each app category, and usage average duration of each app category (refer to Table B.1 for a detailed list of all existing categories).
- **5 dependent variables:** Emotional Stability, Extroversion, Openness, Conscientiousness, and Agreeableness – the Big Five personality traits.

After taking a look at the overall population of this study, a more in-depth analysis is performed. Firstly, the independent variables are considered, secondly the dependent variables, and lastly, the relation between the two.

Correlation analysis is performed several times during EDA. In statistical terms, correlation is considered a method of assessing a possible two-way linear association between two variables [99]. Correlation is measured by the correlation coefficient, which represents the strength of the linear relation between the variables in question. It is a dimensionless quantity that assumes a value from -1 to +1 [100]. A correlation coefficient of zero indicates that no linear relationship exists between two variables, while a correlation coefficient of -1 or +1 indicates that there is a perfect negative or positive linear relationship, respectively. There are two main types of correlation coefficients: Pearson's correlation coefficient and Spearman's correlation coefficient [101], both considered in this study. Even though correlation does not mean causation [102], it may be an interesting starting point for further analysis.

Before moving to the following steps, it should be noted that after performing EDA all data instances are randomly divided into two sets: 75% instances in the training set and 25% instances in the testing set.

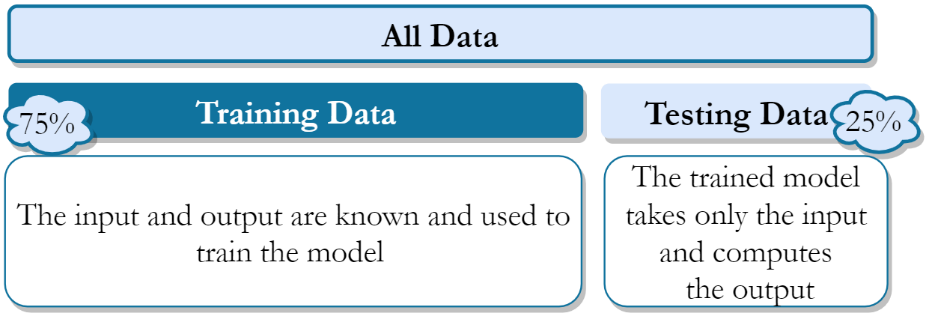


Figure 4.3: Data set split into training and testing set.

All further analysis occurs exclusively on the training set, thereby avoiding any leakage of information about the testing set, which is only used in Section 4.6. Since the frequency and the duration of app usage have different units of measurement, feature scaling is also applied.

4.3 Feature Engineering and Selection

The basic premise of feature engineering and feature selection is that the original data contains information (features) that may not be relevant or may negatively impact the construction of the predictive model [103]. Thus, feature selection is the process of selecting a subset of relevant features from the original ones, while feature engineering creates a new, smaller set of features that still captures most of the useful information [104]. These techniques have many potential benefits, such as: facilitating data visualization and data understanding, reducing training time, and improving the prediction performance [105] since there is an increased risk of models overfitting with an increasing number of features [106].

In this case, there are 55 independent variables. Since this is a reasonable number of features, especially considering the total number of instances, the main intention is to assess the importance of each feature for each personality trait and determine whether or not they are relevant for the final predictive models. Feature engineering and selection is thus handled differently for each trait because different features are likely to impact the five personality traits in different ways.

It should be taken into consideration that there are numerous feature engineering and selection techniques. In order to understand which might be more suited to the data and the model itself, three different techniques are used:

- **Univariate Selection with *SelectPercentile*** : Works by selecting the best features based on univariate statistical tests. This is done according to a percentile of the highest scores which are obtained through *f_classif*. This function computes the ANOVA F-value between label/feature for the classification tasks .
- **Random Forest Classifier and its attribute *feature_importances_***: Returns the feature's importance - the higher, the more important is the feature. The best features are then selected to build the predictive models.
- **Principal Component Analysis (PCA)**: Based on linear dimensionality reduction using Singular Value Decomposition of the data to project it to a lower dimensional space.

The two first techniques are thus considered feature selection, while the last is considered feature engineering.

4.4 Comparing Models On a Metric

As already seen in Chapter 3, previous studies show significant methodological differences. Different tests are used to assess personality trait scores and the behavioral factors that can be taken into consideration are extremely diverse. Therefore, the relationship between personality and behavior is not clear. In order to understand which type of model might be more suited to the data, several classification algorithms are evaluated and compared for each personality trait. Based on the feature selection already implemented, the goal is to find the best algorithm for each trait.

Seven classification algorithms, with different characteristics, are chosen:

- Logistic Regression (LR)
- Linear Discriminant Analysis (LDA)
- K-Neighbors Classifier (KNN)
- Random Forest Classifier (RF)
- Gaussian Naive Bayes (NB)

- Support Vector Classifier (SVM)
- Gradient Boosting Classifier (BOOST)

The chosen algorithms are very diversified, which allows us to understand which are better suited to the problem.

For this initial comparison, Scikit-Learn default values for each model hyperparameters are used. However, these values are not guaranteed to be optimal for this problem. Thus, based on the results obtained, the best models are chosen to undergo hyperparameter tuning in the next section.

K-fold cross-validation is also used to compare the models. This type of cross-validation works by splitting the training data into k-folds where each observation is assigned to an individual group and stays in that group for the entire duration of the process. The model is then trained on k - 1 folds and evaluated on the remaining one. This process repeats itself k times, allowing the models to be trained and evaluated multiple times on different data. Thus, a reliable estimate for the performance of the algorithms is obtained. Figure 4.4 provides a better understanding of how this method works. While the figure demonstrates a five-fold cross-validation, the same principle is applied to any k-fold cross-validation.

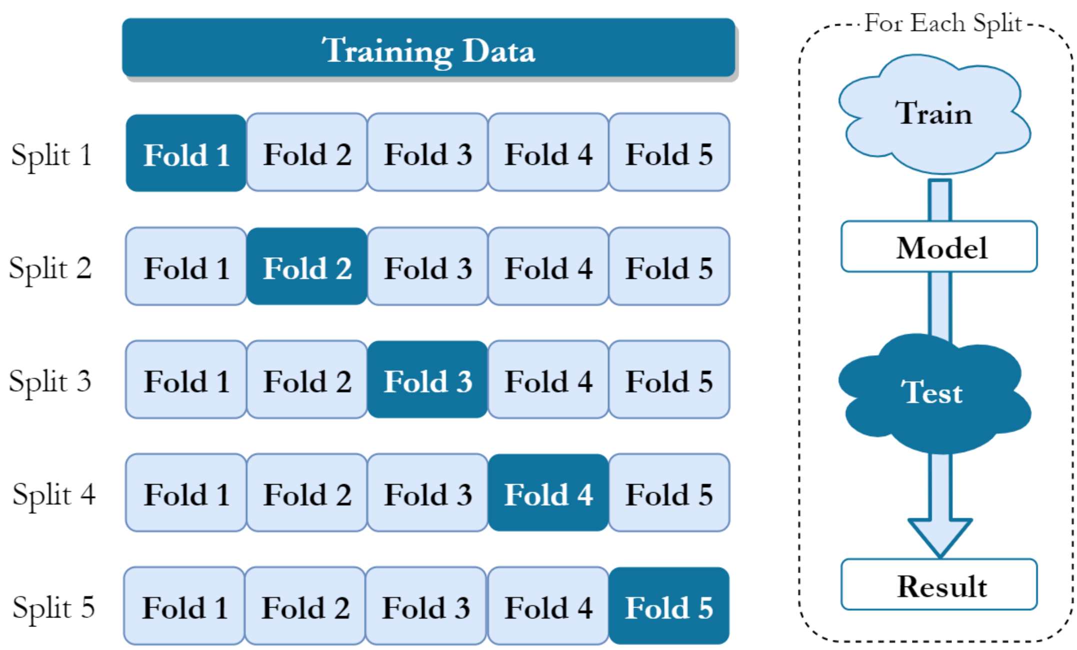


Figure 4.4: Demonstration of a five-fold cross-validation.

4.4.1 Evaluation Metrics

In order to compare the performance of the aforementioned models in the data, distinct evaluation metrics that are often found in the literature are considered: Accuracy, Precision, Recall, and F1-Score.

Confusion Matrix summarizes the performance of the model by providing four important values:

- True Positives (TP): the number of positive outputs correctly predicted.
- False Positives (FP): the number of outputs wrongly predicted as positive.
- True Negatives (TN): the number of negative outputs correctly predicted.
- False Negatives (FN): the number of outputs wrongly predicted as negative.

PREDICTED / ACTUAL	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

Figure 4.5: Example of a confusion matrix.

Figure 4.5 shows an example of a confusion matrix and summarizes how these values are obtained. With them, it is possible to compute the already mentioned metrics as detailed in Equations 4.1, 4.2, 4.3, and 4.4.

Accuracy is the ratio of the number of correct predictions to the total number of samples. In this case, it is a good metric for model comparison because, in both approaches, each class has an approximately equal number of samples.

$$Accuracy = \frac{TP + FP}{TP + TN + FP + FN} \tag{4.1}$$

Precision corresponds to the portion of positive predictions that are actually correct.

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

Recall corresponds to the portion of actual positive outputs that are correctly predicted.

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

F1-Score is the Harmonic Mean between precision and recall and, mathematically, it can be expressed as:

$$\text{F1-Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4.4}$$

4.5 Model Optimization

Machine learning models typically have parameters and hyperparameters:

- **Model Parameters** can be considered what the model learns during training, such as the weights in linear regression [107].
- **Model Hyperparameters** are best thought of settings of a machine learning algorithm that can be manually tuned.

Thus, optimizing a model means finding the best set of hyperparameter values for each particular problem (i.e. the set of hyperparameter values that maximizes the performance of the final predictive models).

The method used is Grid Search Parameter Tuning, which methodically builds and evaluates models for each combination of the given algorithm hyperparameters (refer to Table C.1 for a detailed list of the parameters tested). This is implemented together with k-fold cross-validation for the models obtained from the steps performed in Section 4.4. Ultimately, the chosen hyperparameters for the final classifier are the ones used in the model with the best performance.

Since for each algorithm there is a wide range of possible combinations for the hyperparameter values, this process is time-consuming and involves considerable computer power. That is the reason why it is important to compare models at an earlier stage and choose only the most promising ones for model optimization.

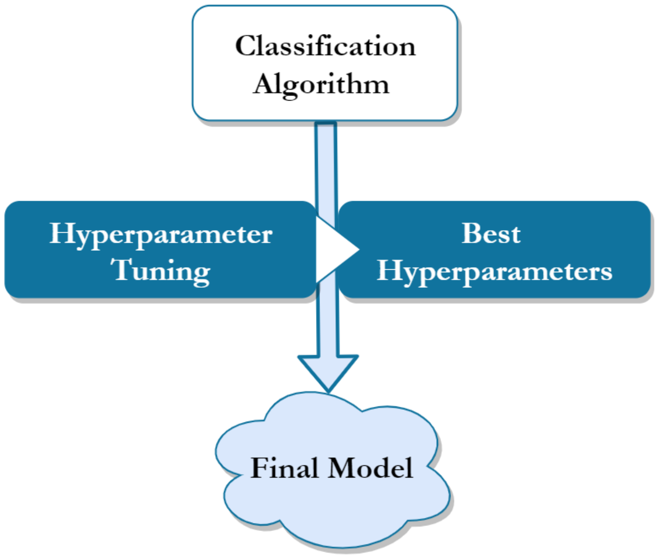


Figure 4.6: Model optimization through hyperparameter tuning.

4.6 Evaluate the Best Model on the Testing Set

At this stage, the best model for each personality trait has already been obtained based on the training set. Thus, the hold-out test set can be seen as the sample of data used to provide an unbiased evaluation of the final model, giving a good insight about how the model behaves in never before seen data. If the performance of the classifiers is good on the testing set, it is more likely that it will also be good if applied to a real-life context.

However, in order to be able to draw any conclusions from the obtained results, it is essential to have a basis for comparison. Thus, a baseline, which is the simplest possible prediction one can make, provides a point of comparison for the more advanced methods developed throughout these experiments [107].

Two of the most commonly used baseline algorithms are:

- Zero Rule Algorithm.
- Random Prediction Algorithm.

The first algorithm, Zero Rule Algorithm, which is also known as a majority class selection, makes predictions for the test set taking into consideration the class value that is most common in the training set. This is extremely useful for unbalanced data sets but, since that is not the case of this problem, a Random Prediction Algorithm is used. This is a very simple algorithm that basically consists in collecting the set of unique output values from the training data, in this case, “high” and “low” for the 1st Experiment and “high”, “medium” and “low” for the 2nd Experiment. Then, a randomly selected output value is chosen from the previously collected set for each row in the test set.

Once the baseline is computed with the Random Prediction Algorithm for each personality trait, there is a meaningful reference point to which the results obtained on the test set can be compared to.

5

Results and Discussion

5.1 Exploratory Data Analysis

The 137 participants present in the data set (87 women and 50 men) are on average 23.6 (S.D.=4.7) years old, with the age ranging from 18 to 50 years old. The obtained sample consisted primarily of students and employees of Ludwig-Maximilians-University of Munich [61] so it is not surprising that the vast majority of the participants (96%) has completed, at least, high school education.

Independent Variables Analysis

The independent variables of this study are gender, usage frequency of each app category, and usage average duration of each app category. By taking a look at app usage in general, it is possible to see that some mobile applications such as Communication, Tools, Browser, and Social apps are more commonly used. Yet, Figure 5.1 reflects how Communication-related apps are much more frequently used than any other. On the other hand, Figure 5.2 shows how other categories like Games, Calls, Entertainment, and Browser apps are used for longer periods of time. In addition, by taking a look at both Figure 5.1 and Figure 5.2, it is clear that the relation between duration and frequency varies from app category to app category.

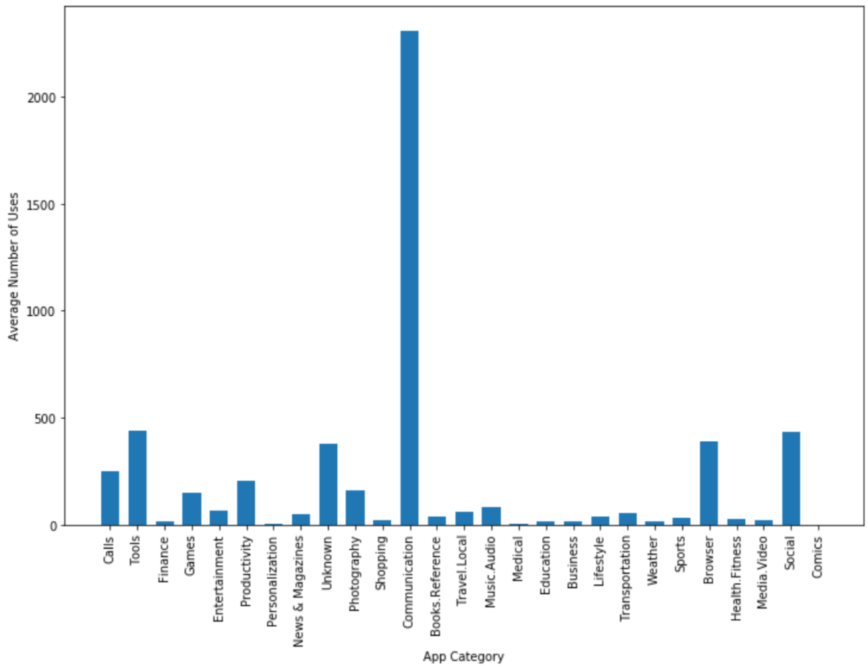


Figure 5.1: Average frequency of use of each mobile app category.

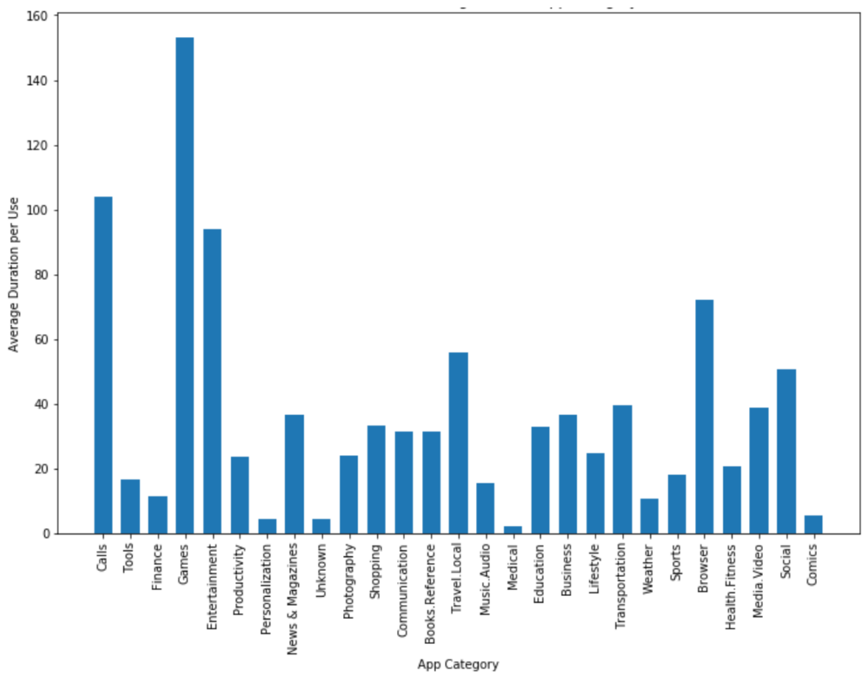


Figure 5.2: Average duration of use of each mobile app category.

Dependent Variables Analysis

The dependent variables of this study are the five personality traits: Emotional Stability, Extroversion, Openness, Conscientiousness, and Agreeableness. Table 5.1 provides a descriptive statistic of these variables. The table includes the mean and standard deviation for the five traits, as well as the minimum and maximum values, and the 25th, 50th, and 75th percentiles. The skewness of the traits distribution is also tabulated. Emotional Stability, Extroversion, and Conscientiousness have a skew very close to zero. On the other hand, Openness and Agreeableness show a slightly higher skew, with the value for Openness being the highest. Nevertheless, all values are between -1 and 1 which means that it can be assumed that all personality traits have a distribution close to normal [108].

Table 5.1: Summary of the central tendency, dispersion and shape of the personality traits distribution.

	Personality Traits				
Metrics	Emotional Stability	Extroversion	Openness	Conscientiousness	Agreeableness
count	137	137	137	137	137
mean	-0.0424	0.0281	0.0141	0.0782	-0.1555
std	0.7044	0.7382	0.7196	0.7714	0.7516
min	-1.9955	-1.9760	-1.8410	-1.6259	-2.1099
25%	-0.4200	-0.4628	-0.4966	-0.4334	-0.6329
50%	-0.0475	-0.0007	-0.1128	0.0590	-0.1890
75%	0.4303	0.4861	0.4416	0.6018	0.2822
max	2.5201	1.8764	2.1176	1.8143	1.8009
skew	0.0741	0.0970	0.5599	0.0110	0.2388

Note: 25%, 50% and 75% correspond to the 25th, 50th, and 75th percentiles, respectively.

Additionally, a Pearson Correlation test is performed to determine how strongly correlated the traits are between them. Table 5.2 summarizes the strength of the linear relationship between each pair of personality traits. All traits are moderately inter-correlated, with Openness and Extroversion reaching the highest value with a correlation of 0.586. Nevertheless, all correlations presented in Table 5.2 are below the selection criteria used in the test for multicollinearity in previous work [61]. These results align with previous studies [92, 96] that report that more open individuals tend to be more extroverted as well. However, it should be taken into consideration that these relations may vary considerably depending on the group of participants being analyzed.

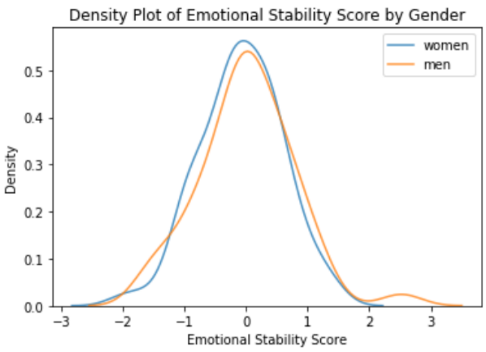
Table 5.2: Correlations between the five personality traits.

Personality Traits	1	2	3	4	5
1. Emotional Stability	1				
2. Extroversion	0.466*	1			
3. Openness	0.330*	0.586*	1		
4. Conscientiousness	0.338*	0.255*	0.294*	1	
5. Agreeableness	0.282*	0.381*	0.424*	0.170	1

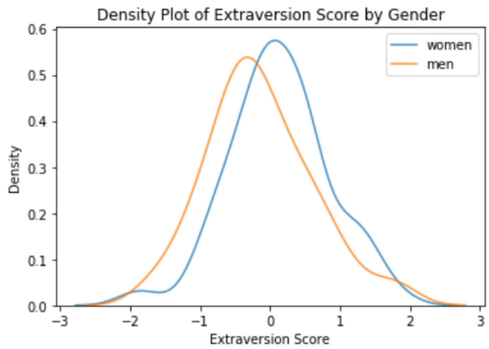
* $p < 0.01$

Relation Between Dependent and Independent Variables

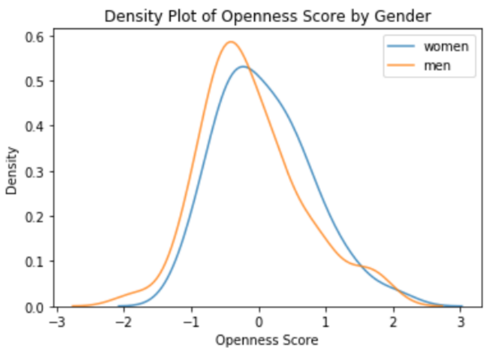
Previous studies [109] mention that men and women often score differently on personality tests and these differences have already been taken into consideration during research similar to this [93, 95]. Therefore, it is appropriate to see whether or not gender affects the personality traits of the participants.



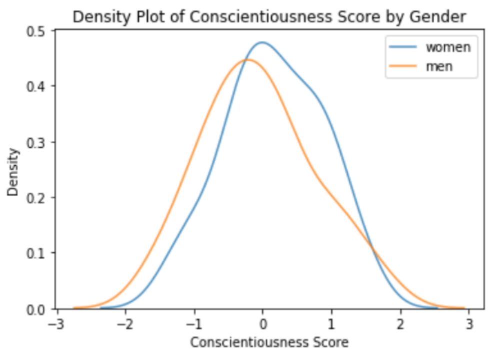
(a) Emotional Stability



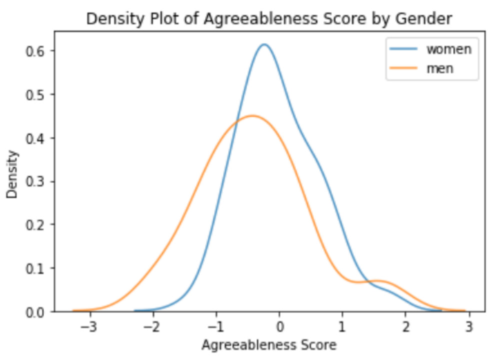
(b) Extroversion



(c) Openness



(d) Conscientiousness



(e) Agreeableness

Figure 5.3: Density Plots of each Personality Traits According to Gender.

Figure 5.3 shows that while some personality traits, like Emotional Stability, don't seem to be too affected by the gender of the participants, others, like Agreeableness, behave in a distinct way. Although the focus of this work is the detection of personality traits from smartphone usage, the user's gender may impact this relation. Thus, this demographic variable is included as an independent variable.

Furthermore, the correlation between all independent variables and the personality traits can be computed. Since there is a deviation from Gaussian distribution in all app usage categories, both for the frequency and the duration, Spearman correlation is used for this analysis [101]. Therefore, the features that show a statistically significant correlation (positive or negative) with each personality trait are considered.

The Emotional Stability trait seems to be positively correlated with a more frequent use of Transportation apps ($\rho = 0.177, p < 0.05$) and shows a negative correlation with the frequency of use of Media and Video apps ($\rho = -0.210, p < 0.05$).

Emotional Stable individuals tend to be calm, relaxed, self-confident and not anxious [47]. Previous research [49] has found that this personality trait tends to be negatively correlated with the use of Photography, Music and Audio, and Penalization apps. According to the authors, people who score low on the Emotional stability trait tend to use these apps due to their "fussy and picky nature" as well as their interest in creative activities. If this is confirmed, a similar explanation may be given about the results obtained regarding the negative correlation between this trait and the frequency of use of Media and Video apps. However, further research is advised before drawing such conclusions.

The second personality trait, Extroversion, shows some positive correlation with the frequency of Calls ($\rho = 0.332, p < 0.001$) and of use of Communication ($\rho = 0.270, p < 0.01$) and Transportation ($\rho = 0.174, p < 0.05$) related apps. The duration of use of Shopping ($\rho = 0.189, p < 0.05$) and Lifestyle ($\rho = 0.175, p < 0.05$) app categories also seems to be positively correlated with this trait.

People who score high on Extroversion tend to value close and interpersonal relationship, and are typically social, outgoing, and talkative [30]. The current findings, that suggest that extroverted individuals tend to perform and receive more calls and use communication apps more often, are highly supported by previous research. Several studies [92, 81] reveal that extroverted individuals were related to an increased use of instant messaging apps, including WhatsApp [56] and give a great importance to the texting function [91]. Both instant messaging apps and texting are comprehended in the Communication category considered in this study.

It was also found that extroverted people tend to receive more calls and spend more time on them [95]. Since there was no positive correlation found between the duration of calls and this personality trait, the results of this study can't validate the fact that extroverted tend to spend more time on calls. However, both agree and support each other regarding Extroversion and an increased number of phone calls.

The communicative characteristic of extroverted individuals is clearly highlighted by the above findings. However, no other research corroborates the positive correlation found between Extroversion and Transportation, Lifestyle, and Shopping apps which indicates that these findings should not be generalized for people outside this data set without performing additional research.

Openness, on the other hand, is a trait that shows negative correlation with the duration of use of Sports related apps ($\rho = -0.203, p < 0.05$) and both the duration ($\rho = -0.193, p < 0.05$) and frequency ($\rho = -0.226, p < 0.01$) of use of Comics apps.

Openness is frequently associated with being imaginative, broad-minded and artistically sensitive. Previous studies related a low score in Openness with higher rates of text messaging. To the best of our knowledge, there are no more links between this trait and any particular behavior regarding smartphone functions and mobile apps. Therefore, at the moment it is hard to compare these findings but they may provide groundings for future analysis.

Conscientiousness reveals a negative correlation with the frequency of use of apps that belong to the category of Travel and Local($\rho = -0.215, p < 0.05$).

Conscientiousness is linked with being careful, thorough, responsible, organized and reliable. Previous studies found that Conscientiousness was linked with a decreased use of Finance and Shopping applications [92], as well as Photography, Music and Audio, and Penalization apps [49]. This trait was also negatively correlated with the use of WhatsApp [56].

These results are scattered but, in general, align with the fact that conscientious people are success driven [31], therefore, less likely to use leisure mobile apps because they may regard them as distracting and unproductive [49]. Notwithstanding, no correlation between this trait and the use of productivity related apps was found.

Lastly, Agreeableness seems to be correlated with a more frequent use of Transportation apps ($\rho = 0.200, p < 0.05$).

People who score high in Agreeableness are courteous, flexible, and tolerant. Previous research shows that people who score higher on this personality trait tend to adopt Penalization apps [49] and report greater importance on calls than on texting

[91]. Once again, these findings are very sparse.

Considering each personality trait individually, there seems to be a wide variety of findings for each of them across different studies. This is probably due to the fact that sociodemographic variables have a stronger relationship with smartphone and mobile application usage than personality [92]. For methodological reasons, most studies collect data from participants belonging to a specific segment of the population leading to data sets that are not representative of all population-levels across different social, cultural and economic backgrounds. The data set used in this study is also an example of this limitation because it takes into consideration only a very specific segment of the German population. Notwithstanding, the personality trait of Extroversion seems to overcome sociodemographic differences and reports solid and consistent results across numerous studies.

On the other hand, on a macro level, the frequency of use of Transportation apps seems to be positively correlated with three different personality traits: Emotional Stability, Extroversion, and Agreeableness. Nevertheless, all other correlations found seem to differ from personality trait to personality trait. Furthermore, during the Independent Variables Analysis, it was seen how Communication-related apps are much more frequently used than any other app category, with Figure 5.1 highlighting this disparity. This might have led one to believe that this app category was more commonly used across all participants, regardless of their personality trait. However, only Extroversion reports a significant correlation with the frequency of use of this app category.

These findings thus suggest that personality traits do have an impact on participants' behavior when it comes to the adoption and use of different mobile applications. With the exception of Transportation-related apps, each personality trait shows very distinct relations with all mobile app categories, which also supports previous research.

Yet, it can not be ignored that there are probably other factors influencing these results. Besides sociodemographic variables, it is important to consider that the use of mobile apps also depends on personal preferences that may, or may not, be linked to personality. In the future, a possible way to prevent the influence of personal taste might be to also capture information from other sources, such as other smartphone sensors, that don't rely as much on personal taste (e.g. Accelerometer, Bluetooth logs, Global Positioning System (GPS) scans, etc.).

5.2 1st Experiment: Binary Classification

In the first experiment, a binary classification task for each of the Big Five personality traits is defined. To do this, three different feature selection and engineering techniques are used as well as seven classification algorithms, all of them already mentioned in Sections 4.3 and 4.4, respectively. In order to understand which technique and classification algorithm is more suited to the data being analyzed, all techniques and algorithms are combined for each personality trait. Table D.1 compiles all the results obtained in this step and allows a thorough comparison of the different approaches. In general, Univariate Selection using *SelectPercentile* shows better results across the different models and for all personality types. Hence, only this technique is considered for further analysis.

Taking into consideration their mean accuracy values and corresponding standard deviation (which can be seen in Table D.1), the following binary models are considered the best for each personality trait and are thus chosen to undergo hyperparameter tuning:

- **Emotional Stability:** Logistic Regression (LR), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (NB), and Support Vector Machine (SVM).
- **Extroversion:** Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Support Vector Classifier (SVM).
- **Openness:** Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), and Gaussian Naive Bayes (NB).
- **Conscientiousness:** Logistic Regression (LR), Random Forest Classifier (RF), Gaussian Naive Bayes (NB), and Gradient Boosting Classifier (BOOST).
- **Agreeableness:** Logistic Regression (LR), Linear Discriminant Analysis (LDA), Random Forest Classifier (RF), and Support Vector Classifier (SVM).

Table D.3 compiles all the obtained results after model optimization for the binary classifiers. This table also contains the values of the parameters that maximize each algorithm performance. The KNN classifier shows the best results for both Emotional Stability and Openness, while Logistic Regression outperforms the other algorithms for Extroversion and Conscientiousness. Lastly, for detecting Agreeableness, a SVM classifier shows the best results. Table 5.3 summarizes this information and shows the accuracy obtained on the training set for each personality trait.

Table 5.3: Best performing algorithm for each personality trait and corresponding accuracy value on the training set - 1st Experiment.

Personality Trait	Best Algorithm	Mean Accurary Value on the Training Set
Emotional Stability	KNN	0.685 (0.079)
Extroversion	LR	0.696 (0.059)
Openness	KNN	0.647 (0.112)
Conscientiousness	LR	0.618 (0.061)
Agreeableness	SVM	0.705 (0.071)

Thus, to evaluate the performance of the aforementioned models on unseen data, these are tested on the hold-out testing set. The obtained results are compiled in Table 5.4, along with the accuracy expected from a random baseline for each personality trait.

Table 5.4: Performance metrics obtained on the testing set for the final classifiers of each personality trait - 1st Experiment.

		Testing Set			
Personality Trait	Random Baseline	Accuracy Value	Precision	Recall	F1-Score
Emotional Stability	50.0%	65.7%	66.0%	66.0%	66.0%
Extroversion	51.4%	65.7%	66.0%	66.0%	65.0%
Openness	48.6%	57.1%	57.0%	57.0%	57.0%
Conscientiousness	51.4%	60.0%	60.0%	60.0%	60.0%
Agreeableness	48.6%	57.1%	65.0%	57.0%	57.0%

Upon comparison of the accuracy values obtained on the training and the testing set, it is possible to see that some personality traits are easier to predict than others. Furthermore, all the classifiers evaluated on the testing set perform above the baseline, proving that personality traits can indeed be predicted from smartphone usage, yet, without a significant degree of certainty. With the baselines being between 48,6% and 51.4%, the best binary classifiers for the five personality traits predict on average 22% better than random.

Interestingly, Emotional Stability and Extroversion are the traits that are best predicted in this 1st Experiment. This is in accordance with previous research that also reports better results for these two personality traits [93]. The authors attribute these findings to the fact that these two traits are the dimensions of personality most directly linked with emotions [110]. In particular, Extroversion is linked with positive emotions whereas the lack of Emotional Stability is associated with negative ones. This may lead to the features picking up the emotional components related to these traits which, ultimately, helps the machine learning models to accurately predict both Emotional Stability and Extroversion.

5.3 2$^{\text{nd}}$ Experiment: Multi-class Classification

The process used for the first Experiment is replicated for the second, which focus on multi-class classification. Table D.2 compiles all the results obtained in this part and allows a thorough comparison of the different approaches. Once again, Univariate Selection using *SelectPercentile* shows the best results. Thus, only this technique is considered for further analysis.

Taking into account their mean accuracy values and corresponding standard deviation (which can be seen in Table D.2), the following 3-label models are considered the best for each personality trait and are thus chosen to undergo hyperparameter tuning:

- **Emotional Stability:** Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (NB), and Support Vector Classifier (SVM).
- **Extroversion:** Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), and Gaussian Naive Bayes (NB).
- **Openness:** Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), and Gaussian Naive Bayes (NB).
- **Conscientiousness:** Linear Discriminant Analysis (LDA) and Random Forest Classifier (RF)
- **Agreeableness:** Linear Discriminant Analysis (LDA), Random Forest Classifier (RF), Gaussian Naive Bayes (NB), and Support Vector Classifier (SVM).

Table D.4 compiles all the obtained results after model optimization for the multi-class classifiers. This table also contains the values of the parameters that maximize each algorithm performance. A SVM classifier shows the best results for Emotional Stability. For the personality trait of Extroversion, both LDA and Naive Bayes report the same mean accuracy and distribution. Naive Bayes is also considered the best algorithm for Openness and Agreeableness. Lastly, a RF classifier outperforms the other algorithms for Conscientiousness. Table 5.5 summarizes this information and shows the accuracy obtained on the training set for each personality trait.

Table 5.5: Best performing algorithm for each personality trait and corresponding accuracy value on the training set - 2nd Experiment.

Personality Trait	Best Algorithm	Mean Accurary Value on the Training Set
Emotional Stability	SVM	0.529 (0.096)
Extroversion	LDA	0.500 (0.057)
	NB	0.500 (0.057)
Openness	NB	0.531 (0.137)
Conscientiousness	RF	0.561 (0.119)
Agreeableness	NB	0.499 (0.106)

Once again, to evaluate the performance of the aforementioned models on unseen data, these are tested on the hold-out testing set. The obtained results are compiled in Table 5.6. Even though the Linear Discriminant Analysis and the Naive Bayes predicted equally well the personality trait of Extroversion on the training set, the LDA doesn't meet the expected performance on unseen data. Therefore, Naive Bayes is considered the best algorithm for predicting Extroversion in this multi-class classification.

Table 5.6: Performance metrics obtained on the testing set for the final classifiers of each personality trait - 2nd Experiment.

		Testing Set			
Personality Trait	Random Baseline	Accuracy Value	Precision	Recall	F1-Score
Emotional Stability	34.3%	37.1%	37.0%	37.0%	36.0%
Extroversion	28.6%	34.3%	33.0%	34.0%	30.0%
Openness	31.4%	37.1%	36.0%	37.0%	36.0%
Conscientiousness	34.3%	37.1%	41.0%	37.0%	36.0%
Agreeableness	31.4%	34.2%	32.0%	34.0%	32.0%

In addition to the results obtained from the best classifiers for each personality trait, the accuracy expected from a random baseline is also provided in Table 5.6.

Similarly to the previous experience, some personality traits are easier to predict than others and all classifiers perform above the baseline. However, in this 2nd Experiment the differences between traits are much more subtle as well as the performance of the models when compared to the baseline. With the last being between 28,6% and 34,3%, the best multi-class classifiers for the five personality traits predict on average 12% better than random.

It was expected that the accuracy value on the training and testing set would be higher for the binary classifiers since they only have two classes to choose from while the multi-class ones had three, instantly decreasing the chances of getting the outcome right. However, the baseline chosen for each experiment takes into consideration this difference. Yet, the multi-class classifiers are only 12% better than random while, in comparison, the binary classifiers perform almost two times better than the previous ones (i.e. 22% better than random). By analyzing each personality trait individually, this trend is confirmed, with the binary classifiers performing better for all the Big Five personality traits.

Taking into consideration the above findings, it should be noted that the lower performance of the multi-class classifiers may be due to lack of enough data. The data set used is not particularly large, thus, by splitting the participants into three classes instead of two, each class is left with far less instances. That makes it harder for the models to learn enough during training to be able to perform well on unseen data. Thus, further comparison between binary and multi-class classification with a larger data set is encouraged for predicting personality traits.

Notwithstanding, in this case, binary classification, performed during the 1st Experiment, is the overall best approach. Not only do the binary classifiers perform better but they also fulfill the primary goal of distinguishing people that belong to opposite poles of each personality trait. By using them, it is possible to differentiate introverts from extroverts, compassionate from detached individuals, nervous from confident, curious from cautious, and organized from careless better than chance. This alone would greatly benefit IPAs' capability of adapting to their users, that in turn would lead to a higher value of the support provided as well as a higher acceptance among users. With further development, these models can also be useful to many other fields.

Even though the predictive models obtained don't show exceptional results, this work contributes to the state of the art with a detailed machine learning approach, which can be easily escalated and used for studies that have access to more data. However, it is often difficult to gather participants for this kind of research. In this day and age when computing power is less and less a problem, data collection is one of the major obstacles that still needs to be overcome in this field.

6

Conclusions

The main goal of this study was to determine whether it is possible to predict someone's personality traits based on their activity on their smartphone, specifically their mobile app usage. For this, it was important to understand if people with different personality traits tend to use different apps. In addition, a machine learning approach was chosen and two different classification types were implemented – binary and multi-class.

This study starts by confirming previous findings claiming that personality traits have an impact on the usage of different mobile apps. The most notable discovery to take in consideration is that the personality trait of Extroversion seems to be related to a higher number of phone calls and a frequent use of Communication apps. This clearly aligns with psychological assessments that describe extroverted individuals as social and communicative. Furthermore, this is supported by a significant amount of research that reached the same conclusions, leading us to believe that these findings can be generalized with confidence.

Moreover, it was found that Extroversion also shows a positive correlation with the frequency of use of Transportation apps and the duration of use of Shopping and Lifestyle apps. Emotional Stability shows a positive correlation with the frequency of use of Transportation apps as well, while also having a negative correlation with the frequency of use Media and Video apps. Openness is characterized by a negative correlation with the duration of use of Sports related apps and both the duration and frequency of use of Comics apps. Conscientiousness reveals a negative correlation with the frequency of use of Travel and Local apps. Finally, Agreeableness seems to be correlated with a more frequent use of Transportation apps. These findings deserve further investigation in order to be able to draw additional conclusions.

In the following stage, a supervised learning approach was developed. In order to predict the personality traits of the participants, there were first developed five binary classifiers and then five multi-class classifiers. During the first experiment, that focused on binary classification, Emotional Stability and Extroversion were the traits that were most easily predicted, both with an accuracy of 65,7% on unseen data.

All the ten final models obtained performed above random. However, binary classifiers seem to be more adequate for this problem. Not only is the accuracy value on the training and testing set of the binary classifiers higher but also is their performance when compared to the baseline. The binary classifiers are almost 2 times better than the multi-class ones when compared to random. To the best of our knowledge, this is the first work to compare a binary and multi-class classification for personality detection.

The obtained binary classifiers show room for improvement. Yet, they allow the distinction between people who score higher or lower in each personality trait, making it possible to deduce someone's characteristics without really knowing them. This alone would greatly benefit IPAs capability of adapting to their users, both by leading to a higher value of the support provided and a higher acceptance among users. With further development, these models can be useful to many other fields like advertising, human resources management, mental health monitoring, among many others.

Notwithstanding, the obtained results should be viewed in light of the limitations of this study. Firstly, and similarly to previous studies, the data set was obtained from a very specific segment of the population, which may not be representative of all population levels and sociodemographics, therefore influencing the results.

Secondly, nowadays it is possible to use a wide range of smartphone components as data sources. Information can be collected from a smartphone's accelerometer, Bluetooth radio, GPS, Light sensor, Microphone, Wi-Fi scans, Cameras, Phone use logs, and App use logs [60]. However, the data set used in this work takes only into consideration the last three, not allowing the tracking of many aspects of the participants' daily lives. As an example, it is not possible to record any information about the physical movement of the individuals, which can be tracked by the accelerometer, GPS, or even Wi-Fi scans. In addition, app usage alone may depend a lot on personal preferences that are not related to personality. By adding other variables, that don't depend as much on it, the influence of personal preferences may be mitigated.

Lastly, it should be taken into consideration the ethical and privacy implications of this kind of research. Studies that imply a systematic collection of personal information often bring up privacy concerns. What is the trade-off between the benefits of this personal data acquisition and the ethical concerns that arise from it? It is still unclear since this is a fairly recent problem resultant of our current ability to automatically collect and store considerable amounts of data. Some [111] think that the entities collecting this data should always focus on participants' safety as well as obtaining their informed consent and assuring the confidentiality of all the information. However, with smartphones being the most personal devices people own, it is still common for participants in this kind of studies to be hesitant about sharing so much intimate information. If people have concerns about sharing personal information to the interest of science, one can only imagine how much greater these concerns will be if this data is collected by a service provider. To what extent will the general population be willing to share personal information in exchange for better services? Only time will tell.

6.1 Future Work

Taking into consideration the limitations of this and other studies, it is clear that there is considerable room for improvement in the field of personality detection through smartphone usage. To further advance this field of research, an investment should be made in data collection, both on the amount and diversity of samples gathered as well as the variety of features. After all, data lays the foundation for all additional analysis.

It is important to consider that sociodemographic variables are strongly related to smartphone usage and try to collect information from participants from different social, cultural, and economic backgrounds. This together with an increase in the amount of data collected can lead to more meaningful research and conclusions that may be generalized for people outside the data sets.

In addition, it might be interesting to broaden the features collected and take full advantage of all data sources present in today's smartphones. Taking the previous example, the physical movement has proven to be important for the study of happiness and mental health [12, 112, 113] as well as students' academic performance [114, 115]. Recent studies [96] have already begun to also relate physical movement to personality, which proves that this is an interesting approach that deserves further analysis. Since physical movement seems to be related to happiness and mental health, it is possible that its study may be of significant interest for the personality trait of Emotional stability.

Furthermore, it may be interesting to combine all the information gathered from smartphone sensors with information collected from public social media platforms, which have proven to be related with the personality of their users. The analysis of text massages may also be considered after weighing the pros and cons of its privacy implications. By diversifying the sources of information on each individual, it may be possible to obtain a more comprehensive profile of the user's personality.

All these changes in the data collected can help machine learning models to perform better on unseen data therefore significantly improving their usefulness in real life, presumably the context of ultimate interest.

Finally, other methodological approaches can be taken to analyze the same problem. While this work focuses on supervised learning, it can also be interesting to try a semi-supervised learning approach that might achieve a similar level of performance with a smaller amount of data. An alternative approach would also be to try Deep Learning which concerns artificial neural networks, inspired by the structure and function of the brain. This is a fairly recent but powerful technology that may outperform the classic machine learning techniques.

Bibliography

[1] A. Vital, "Big five personality traits." Available at: `https://blog.adioma.com/5-personality-traits-infographic/`. Online; accessed 18 June 2019.

[2] Mathworks, "What is machine learning? — how it works, techniques applications." Available at: `https://www.mathworks.com/discovery/machine-learning.html`. Online; accessed 22 May 2019.

[3] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 273–291, 2014.

[4] A. Reis, D. Paulino, H. Paredes, and J. Barroso, "Using intelligent personal assistants to strengthen the elderlies' social bonds," in *International Conference on Universal Access in Human-Computer Interaction*, pp. 593–602, Springer, 2017.

[5] G. Czibula, A.-M. Guran, I. G. Czibula, and G. S. Cojocar, "Ipa-an intelligent personal assistant agent for task performance support," in *2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing*, pp. 31–34, IEEE, 2009.

[6] J. Santos, J. J. Rodrigues, B. M. Silva, J. Casal, K. Saleem, and V. Denisov, "An iot-based mobile gateway for intelligent personal assistants on mobile health environments," *Journal of Network and Computer Applications*, vol. 71, pp. 194–204, 2016.

[7] M. L. Walters, K. Dautenhahn, R. Te Boekhorst, K. L. Koay, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, and I. Werry, "The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pp. 347–352, IEEE, 2005.

[8] A. M. Kaplan and M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media," *Business horizons*, vol. 53, no. 1, pp. 59–68, 2010.

[9] K. M. Camisa, M. A. Bockbrader, P. Lysaker, L. L. Rae, C. A. Brenner, and B. F. O'Donnell, "Personality traits in schizophrenia and related personality disorders," *Psychiatry Research*, vol. 133, no. 1, pp. 23–33, 2005.

[10] R. Ferdous, V. Osmani, and O. Mayora, "Smartphone app usage as a predictor of perceived stress levels at workplace," in *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pp. 225–228, IEEE, 2015.

[11] N. F. BinDhim, A. M. Shaman, L. Trevena, M. H. Basyouni, L. G. Pont, and T. M. Alhawassi, "Depression screening via a smartphone app: Cross-country user characteristics and feasibility," *Journal of the American Medical Informatics Association*, vol. 22, no. 1, pp. 29–34, 2014.

[12] L. Canzian and M. Musolesi, "Trajectories of depression: Unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pp. 1293–1304, ACM, 2015.

[13] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, and D. C. Mohr, "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study," *Journal of medical Internet research*, vol. 17, no. 7, 2015.

[14] A. Grunerbl, A. Muaremi, V. Osmani, G. Bahle, S. Ohler, G. Troster, O. Mayora, C. Haring, and P. Lukowicz, "Smartphone-based recognition of states and state changes in bipolar disorder patients," *IEEE Journal of Biomedical and Health Informatics*, vol. 1, no. 19, pp. 140–148, 2015.

[15] M. Matthews, S. Abdullah, E. Murnane, S. Voida, T. Choudhury, G. Gay, and E. Frank, "Development and evaluation of a smartphone-based measure of social rhythms for bipolar disorder," *Assessment*, vol. 23, no. 4, pp. 472–483, 2016.

[16] Y.-H. Lin, Y.-C. Lin, Y.-H. Lee, P.-H. Lin, S.-H. Lin, L.-R. Chang, H.-W. Tseng, L.-Y. Yen, C. C. Yang, and T. B. Kuo, "Time distortion associated with smartphone addiction: Identifying smartphone addiction via a mobile application (app)," *Journal of psychiatric research*, vol. 65, pp. 139–145, 2015.

[17] C. Stachl, S. Hilbert, J.-Q. Au, D. Buschek, A. De Luca, B. Bischl, H. Hussmann, and M. Bühner, "Personality traits predict smartphone usage." Available at: `https://osf.io/b45vq/?view_only=25d00f6868fe44d6bec2177e810223c3`, 2017. Online; accessed 27 January 2019.

[18] G. Matthews, I. J. Deary, and M. C. Whiteman, *Personality traits*. Cambridge University Press, 2009.

[19] M. E. Mikulincer, P. R. Shaver, M. Cooper, and R. J. Larsen, *APA handbook of personality and social psychology, Volume 4: Personality processes and individual differences.* American Psychological Association, 2015.

[20] L. R. Goldberg, "The structure of phenotypic personality traits.," *American psychologist*, vol. 48, no. 1, p. 26, 1993.

[21] G. W. Allport, "Personality: A psychological interpretation.," 1937.

[22] R. B. Cattell and H. E. P. Cattell, "Personality structure and the new fifth edition of the 16pf," *Educational and Psychological Measurement*, vol. 55, no. 6, pp. 926–937, 1995.

[23] H. J. Eysenck, *Dimensions of personality*, vol. 5. Transaction Publishers, 1950.

[24] K. C. Briggs, *Myers-Briggs type indicator.* Consulting Psychologists Press Palo Alto, CA, 1976.

[25] J. M. Digman, "Personality structure: Emergence of the five-factor model," *Annual review of psychology*, vol. 41, no. 1, pp. 417–440, 1990.

[26] S. Matz, Y. W. F. Chan, and M. Kosinski, "Models of personality," in *Emotions and Personality in Personalized Services*, pp. 35–54, Springer, 2016.

[27] O. P. John, S. Srivastava, *et al.*, "The big five trait taxonomy: History, measurement, and theoretical perspectives," *Handbook of personality: Theory and research*, vol. 2, no. 1999, pp. 102–138, 1999.

[28] W. S. Dunn, M. K. Mount, M. R. Barrick, and D. S. Ones, "Relative importance of personality and general mental ability in managers' judgments of applicant qualifications.," *Journal of applied psychology*, vol. 80, no. 4, p. 500, 1995.

[29] M. R. Barrick and M. K. Mount, “The big five personality dimensions and job performance: a meta-analysis,” *Personnel psychology*, vol. 44, no. 1, pp. 1–26, 1991.

[30] D. Watson and L. A. Clark, “Extraversion and its positive emotional core,” in *Handbook of personality psychology*, pp. 767–793, Elsevier, 1997.

[31] J. Uffen, N. Kaemmerer, and M. H. Breitner, “Personality traits and cognitive determinants–an empirical investigation of the use of smartphone security measures,” *Journal of Information Security 4 (2013), Nr. 4*, vol. 4, no. 4, pp. 203–212, 2013.

[32] M. R. Barrick and M. K. Mount, “Select on conscientiousness and emotional stability,” *Handbook of principles of organizational behavior*, vol. 15, p. 28, 2000.

[33] C. G. DeYoung, L. C. Quilty, and J. B. Peterson, “Between facets and domains: 10 aspects of the big five.,” *Journal of personality and social psychology*, vol. 93, no. 5, p. 880, 2007.

[34] P. T. Costa Jr and R. R. McCrae, *Personality in adulthood: A five-factor theory perspective.* Guilford Press, 2013.

[35] P. T. Costa Jr and R. R. McCrae, “Domains and facets: Hierarchical personality assessment using the revised neo personality inventory,” *Journal of personality assessment*, vol. 64, no. 1, pp. 21–50, 1995.

[36] T. J. Bouchard Jr, “Genetic influence on human psychological traits: A survey,” *Current Directions in Psychological Science*, vol. 13, no. 4, pp. 148–151, 2004.

[37] D. Goleman, “Major personality study finds that traits are mostly inherited,” *The New York Times*, 1986. Available at: `https://www.nytimes.com/1986/12/02/science/major-personality-study-finds-that-traits-are-mostly-inherited.html`. Online; accessed 10 July 2019.

[38] J. Clear, *Atomic habits: An easy & proven way to build good habits & break bad ones.* Random House Business Books, 2018.

[39] S. E. Groppe, A. Gossen, L. Rademacher, A. Hahn, L. Westphal, G. Gründer, and K. N. Spreckelmeyer, “Oxytocin influences processing of socially relevant

cues in the ventral tegmental area of the human brain," *Biological psychiatry*, vol. 74, no. 3, pp. 172–179, 2013.

[40] J. A. Bartz, J. Zaki, N. Bolger, and K. N. Ochsner, "Social effects of oxytocin in humans: context and person matter," *Trends in cognitive sciences*, vol. 15, no. 7, pp. 301–309, 2011.

[41] J. Ormel, A. Bastiaansen, H. Riese, E. H. Bos, M. Servaas, M. Ellenbogen, J. G. Rosmalen, and A. Aleman, "The biological and psychological basis of neuroticism: current status and future directions," *Neuroscience & Biobehavioral Reviews*, vol. 37, no. 1, pp. 59–72, 2013.

[42] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german," *Journal of research in Personality*, vol. 41, no. 1, pp. 203–212, 2007.

[43] M. Arendasy, "Bfsi big five structure inventory." Available at: `https://www.schuhfried.com/test/BFSI`. Online; accessed 19 February 2019.

[44] P. T. Costa and R. R. McCrae, "The neo personality inventory," 1985.

[45] P. T. Costa and R. R. Mac Crae, *Neo Personality Inventory-Revised (NEO PI-R)*. Psychological Assessment Resources Odessa, FL, 1992.

[46] P. T. Costa and R. R. McCrae, "The revised neo personality inventory (neo pi-r)," *The SAGE handbook of personality theory and assessment*, vol. 2, no. 2, pp. 179–198, 2008.

[47] S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr, "A very brief measure of the big-five personality domains," *Journal of Research in personality*, vol. 37, no. 6, pp. 504–528, 2003.

[48] M. McGue, S. Bacon, and D. T. Lykken, "Personality stability and change in early adulthood: A behavioral genetic analysis.," *Developmental psychology*, vol. 29, no. 1, p. 96, 1993.

[49] R. Xu, R. M. Frey, E. Fleisch, and A. Ilic, "Understanding the impact of personality traits on mobile app adoption–insights from a large-scale field study," *Computers in Human Behavior*, vol. 62, pp. 244–256, 2016.

[50] I. J. Deary, A. Weiss, and G. D. Batty, "Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health in-

equalities," *Psychological science in the public interest*, vol. 11, no. 2, pp. 53–79, 2010.

[51] M. J. Shanahan, P. L. Hill, B. W. Roberts, J. Eccles, and H. S. Friedman, "Conscientiousness, health, and aging: the life course of personality model.," *Developmental Psychology*, vol. 50, no. 5, p. 1407, 2014.

[52] J. L. Lastovicka and E. A. Joachimsthaler, "Improving the detection of personality-behavior relationships in consumer research," *Journal of Consumer Research*, vol. 14, no. 4, pp. 583–587, 1988.

[53] C. Montag and J. D. Elhai, "A new agenda for personality psychology in the digital age?," *Personality and Individual Differences*, vol. 147, pp. 128–134, 2019.

[54] P. J. Corr and G. Matthews, *The Cambridge Handbook of Personality Psychology.* Cambridge University Press Cambridge, UK:, 2009.

[55] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 74–79, 2017.

[56] C. Montag, K. Błaszkiewicz, R. Sariyska, B. Lachmann, I. Andone, B. Trendafilov, M. Eibes, and A. Markowetz, "Smartphone usage in the 21st century: who is active on whatsapp?," *BMC research notes*, vol. 8, no. 1, p. 331, 2015.

[57] J. Poushter *et al.*, "Smartphone ownership and internet usage continues to climb in emerging economies," *Pew Research Center*, vol. 22, pp. 1–44, 2016.

[58] P. E. Ross, T. S. Romero, W. D. Jones, A. Bleicher, J. Calamia, J. Middleton, R. Stevenson, S. K. Moore, S. Upson, D. Schneider, *et al.*, "Top 11 technologies of the decade," *IEEE Spectrum*, vol. 48, no. 1, pp. 27–63, 2011.

[59] S. Butt and J. G. Phillips, "Personality and self reported mobile phone use," *Computers in Human Behavior*, vol. 24, no. 2, pp. 346–360, 2008.

[60] G. M. Harari, S. R. Müller, M. S. Aung, and P. J. Rentfrow, "Smartphone sensing methods for studying behavior in everyday life," *Current Opinion in Behavioral Sciences*, vol. 18, pp. 83–90, 2017.

[61] C. Stachl, S. Hilbert, J.-Q. Au, D. Buschek, A. De Luca, B. Bischl, H. Hussmann, and M. Bühner, "Personality traits predict smartphone usage," *European Journal of Personality*, vol. 31, no. 6, pp. 701–722, 2017.

[62] W. Wang, G. M. Harari, R. Wang, S. R. Müller, S. Mirjafari, K. Masaba, and A. T. Campbell, “Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, p. 141, 2018.

[63] M. D. Back, S. C. Schmukle, and B. Egloff, “Predicting actual behavior from the explicit and implicit self-concept of personality.,” *Journal of personality and social psychology*, vol. 97, no. 3, p. 533, 2009.

[64] Techopedia, “Mobile application (mobile app).” Available at: `https://www.techopedia.com/definition/2953/mobile-application-mobile-app`. Online; accessed 19 February 2019.

[65] Statista, “Number of apps available in leading app stores as of 3rd quarter 2018.” Available at: `https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/`, 2018. Online; accessed 19 February 2019.

[66] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, “Diversity in smartphone usage,” in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pp. 179–194, ACM, 2010.

[67] P. Welke, I. Andone, K. Blaszkiewicz, and A. Markowetz, “Differentiating smartphone users by app usage,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 519–523, ACM, 2016.

[68] T. Ryan and S. Xenos, “Who uses facebook? an investigation into the relationship between the big five, shyness, narcissism, loneliness, and facebook usage,” *Computers in human behavior*, vol. 27, no. 5, pp. 1658–1664, 2011.

[69] S. Seneviratne, A. Seneviratne, P. Mohapatra, and A. Mahanti, “Predicting user traits from a snapshot of apps installed on a smartphone,” *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 18, no. 2, pp. 1–8, 2014.

[70] J. Shen, O. Brdiczka, and J. Liu, “A study of facebook behavior: What does it tell about your neuroticism and extraversion?,” *Computers in Human Behavior*, vol. 45, pp. 32–38, 2015.

[71] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Pearson Education, 3 ed., 2010.

[72] A. Kaplan and M. Haenlein, “Siri, siri, in my hand: Who’s the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence,” *Business Horizons*, vol. 62, no. 1, pp. 15–25, 2019.

[73] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems.* ” O’Reilly Media, Inc.”, 2017.

[74] J. Bhattacharjee, “Key machine learning definitions.” Available at: `https://medium.com/technology-nineleaps/some-key-machine-learning-definitions-b524eb6cb48`. Online; accessed 10 July 2019.

[75] A. Ng, “Machine learning.” Available at: `https://www.coursera.org/learn/machine-learning`. Online; accessed 22 May 2019.

[76] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach.* Prentice-Hall, 1 ed., 1995.

[77] J. Brownlee, “Supervised and unsupervised machine learning algorithms.” Available at: `https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/`. Online; accessed 22 May 2019.

[78] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, “Scikit-learn: Machine learning in python,” *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[79] D. Kuhlman, *A python book: Beginning python, advanced python, and python exercises.* Dave Kuhlman Lutz, 2009.

[80] A. Luashchuk, “8 reasons why python is good for artificial intelligence and machine learning.” Available at: `https://djangostars.com/blog/why-python-is-good-for-artificial-intelligence-and-machine-learning/`. Online; accessed 23 May 2019.

[81] T. Correa, A. W. Hinsley, and H. G. De Zuniga, “Who interacts on the web?: The intersection of users’ personality and social media use,” *Computers in human behavior*, vol. 26, no. 2, pp. 247–253, 2010.

[82] T. D. Yeo, "Modeling personality influences on youtube usage," in *Fourth International AAAI Conference on Weblogs and Social Media*, 2010.

[83] B. Ferwerda, M. Schedl, and M. Tkalcic, "Predicting personality traits with instagram pictures," in *Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015*, pp. 7–10, ACM, 2015.

[84] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our twitter profiles, our selves: Predicting personality with twitter," in *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pp. 180–185, IEEE, 2011.

[85] L. Qiu, H. Lin, J. Ramsay, and F. Yang, "You are what you tweet: Personality expression and perception on twitter," *Journal of Research in Personality*, vol. 46, no. 6, pp. 710–718, 2012.

[86] F. Liu, J. Perez, and S. Nowson, "A language-independent and compositional model for personality trait recognition from short texts," *arXiv preprint arXiv:1610.04345*, 2016.

[87] G. Carducci, G. Rizzo, D. Monti, E. Palumbo, and M. Morisio, "Twitpersonality: Computing personality traits from tweets using word embeddings and supervised learning," *Information*, vol. 9, no. 5, p. 127, 2018.

[88] M. Skowron, M. Tkalčič, B. Ferwerda, and M. Schedl, "Fusing social media cues: personality prediction from twitter and instagram," in *Proceedings of the 25th international conference companion on world wide web*, pp. 107–108, International World Wide Web Conferences Steering Committee, 2016.

[89] S. Argamon, S. Dhawle, M. Koppel, and J. W. Pennebaker, "Lexical predictors of personality type," in *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America*, pp. 1–16, 2005.

[90] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *Journal of artificial intelligence research*, vol. 30, pp. 457–500, 2007.

[91] W. Lane and C. Manner, "The impact of personality traits on smartphone ownership and use," *International Journal of Business and Social Science*, vol. 2, no. 17, 2011.

[92] Y. Kim, D. A. Briley, and M. G. Ocepek, "Differential innovation of smartphone and application use by sociodemographics and personality," *Computers in Human Behavior*, vol. 44, pp. 141–147, 2015.

[93] Y.-A. de Montjoye, J. Quoidbach, F. Robic, and A. S. Pentland, "Predicting personality using novel mobile phone-based metrics," in *International conference on social computing, behavioral-cultural modeling, and prediction*, pp. 48–55, Springer, 2013.

[94] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Who's who with big-five: Analyzing and classifying personality traits with smartphones," in *2011 15th Annual international symposium on wearable computers*, pp. 29–36, IEEE, 2011.

[95] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Mining large-scale smartphone data for personality studies," *Personal and Ubiquitous Computing*, vol. 17, no. 3, pp. 433–450, 2013.

[96] A. C. Viana, A. Di Luzio, K. Jaffrès-Runser, A. Mei, and J. Stefa, "Accurately inferring personality traits from the use of mobile technology," 2018.

[97] R. C. MacCallum, S. Zhang, K. J. Preacher, and D. D. Rucker, "On the practice of dichotomization of quantitative variables.," *Psychological methods*, vol. 7, no. 1, p. 19, 2002.

[98] P. Patil, "What is exploratory data analysis?." Available at: `https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15`. Online; accessed 30 May 2019.

[99] D. G. Altman, *Practical statistics for medical research.* CRC press, 1990.

[100] T. D. V. Swinscow, M. J. Campbell, *et al.*, *Statistics at square one.* Bmj London, 2002.

[101] M. M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Medical Journal*, vol. 24, no. 3, pp. 69–71, 2012.

[102] J. Lee Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician*, vol. 42, no. 1, pp. 59–66, 1988.

[103] M. L. Bermingham, R. Pong-Wong, A. Spiliopoulou, C. Hayward, I. Rudan, H. Campbell, A. F. Wright, J. F. Wilson, F. Agakov, P. Navarro, *et al.*, "Application of high-dimensional feature selection: evaluation for genomic prediction in man," *Scientific reports*, vol. 5, p. 10312, 2015.

[104] A. Desarda, "Getting data ready for modelling: Feature engineering, feature selection, dimension reduction." Available at: `https://towardsdatascience.com/getting-data-ready-for-modelling-feature-engineering-feature-selection-dimension-reduction-39dfa267b95a`. Online; accessed 29 May 2019.

[105] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.

[106] S. Asaithambi, "Why, how and when to apply feature selection." Available at: `https://towardsdatascience.com/why-how-and-when-to-apply-feature-selection-e9c69adfabf2`. Online; accessed 30 May 2019.

[107] J. Brownlee, "Machine learning mastery with python," 2016.

[108] GoodDataDocumentation, "Normality testing - skewness and kurtosis." Available at: `https://help.gooddata.com/doc/en/reporting-and-dashboards/maql-analytical-query-language/maql-expression-reference/aggregation-functions/statistical-functions/predictive-statistical-use-cases/normality-testing-skewness-and-kurtosis`. Online; accessed 26 June 2019.

[109] A. Feingold, "Gender differences in personality: A meta-analysis," *Psychological bulletin*, vol. 116, no. 3, p. 429, 1994.

[110] A. Gomez and R. Gomez, "Personality traits of the behavioural approach and inhibition systems: Associations with processing of emotional stimuli," *Personality and Individual Differences*, vol. 32, no. 8, pp. 1299–1316, 2002.

[111] M. B. Kapp, "Ethical and legal issues in research involving human subjects: do you want a piece of me?," *Journal of clinical pathology*, vol. 59, no. 4, pp. 335–339, 2006.

[112] N. Lathia, G. M. Sandstrom, C. Mascolo, and P. J. Rentfrow, "Happier people live more active lives: Using smartphones to link happiness and physical activity," *PloS one*, vol. 12, no. 1, p. e0160589, 2017.

[113] S. Abdullah, M. Matthews, E. Frank, G. Doherty, G. Gay, and T. Choudhury, "Automatic detection of social rhythms in bipolar disorder," *Journal of the American Medical Informatics Association*, vol. 23, no. 3, pp. 538–543, 2016.

[114] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "Studentlife: assessing mental health, academic

performance and behavioral trends of college students using smartphones," in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pp. 3–14, ACM, 2014.

[115] J. Fernandes, D. Raposo, S. Sinche, N. Armando, J. S. Silva, A. Rodrigues, L. Macedo, H. G. Oliveira, and F. Boavida, "A human-in-the-loop cyber-physical approach for students performance assessment," in *Proceedings of the Fourth International Workshop on Social Sensing*, pp. 36–42, ACM, 2019.

Appendices

A

Personality

Table A.1: Personality Factors and corresponding Personality Facets

Factor	Facet
Extroversion	Friendliness (E1) Sociableness (E2) Assertiveness (E3) Dynamism (E4) Adventurousness (E5) Cheerfulness (E6)
Emotional Stability	Carefreeness (ES1) Equanimity (ES2) Positive mood (ES3) Self-consciousness (ES4) Self-control (ES5) Emotional Robustness (ES6)
Agreeableness	Willingness to Trust (A1) Genuineness (A2) Helpfulness (A3) Obligingness (A4) Modesty (A5) Good-naturedness (A6)
Conscientiousness	Competence (C1) Love of Order (C2) Sense of Duty (C3) Ambition (C4) Discipline (C5) Caution (C6)
Openness	Openness to Imagination (O1) Openness to Aesthetics (O2) Openness to Feelings (O3) Openness to Actions (O4) Openness to Ideas (O5) Openness to the Value and Norm System (O6)

B

Data Set

Table B.1: Descriptive information of the App Categories taken into consideration

Category	Number Apps	Number Users	Most Frequently Used Apps
Communication	62	137	WhatsApp, Mail, Contacts, Dialer, SMS/MMS
Social	80	125	Facebook, Instagram, Snapchat, Twitter, Weibo
Tools	225	137	Google Search, Clock, Google Play Store, Calculator, S Voice
Browser	6	136	Internet, Firefox, Opera, Dolphin Browser, UC Browser
Calls	1	137	Phone
Productivity	134	137	Settings, S Planner, Calendar, ColorNote, Google Drive
Photography	47	137	Gallery, Camera, SnapApp, Album, PicsArt
Games	229	100	Clash of Clans, Quizduell, Candy Crush Saga, Farm Heroes Saga, Trials Frontier
Music & Games	78	134	Spotify, Music Player, Google Play Music, MP3-Player, SoundCloud
Entertainment	98	131	YouTube, 9GAG, PlayerPro, appinio, PS4-Magazin
Travel & Local	85	134	Maps, MVV Companion, TripAdvisor, BlaBlaCar, Airbnb
Transportation	40	110	MVG Fahrinfo, DB Navigator, , MeinFernbus, Uber
News & Magazines	52	118	FOCUS Online, reddit sync, SPIEGEL ONLINE, Flipboard, SZ.de
Lifestyle	72	72	Tinder, Sleep, Chefkoch, eBay Kleinanzeigen, PAYBACK
Sports	38	33	kicker, Comunio, Kicktipp, Score!, Sportschau
Book & Reference	71	123	Munpia, dict.cc plus, dict.cc, Wikipedia, LEO
Health & Fitness	59	60	SleepBot, Strava, Fitbit, Freeletics, MyFitnessPal
Media & Video	47	118	Video-Player, Google Play Movies, VLC, Video anzeigen, ZDF
Shopping	45	68	eBay, mydealz, Amazon, brands4friends, Shpock
Business	40	108	Eigene Dateien, AnyConnect, POLARIS Oce Viewer 5, Polaris Viewer 4.1, OceSuite
Education	67	54	UnlockYourBrain, AnkiDroid, TUM Campus App, Duolingo, Web Opac
Finance	34	39	Sparkasse, Banking 4A, Wstenrot, YNAB, Banking
Weather	17	74	Weather, wetter.com, WetterOnline, WetterApp, Wetter-Widget
Medical	10	17	Lady Pill Reminder, PillReminder, Pillreminder, iPhysikum, Remember Your Pill
Personalization	14	22	Dokumente, Backgrounds, Zedge, Flatastico, HD Widgets
Comics	6	6	xkcd Browser, NICHTLUSTIG, Marvel Unlimited, xkcdViewer, xkcd - Now

Note: "Num Apps" is the total number of apps in the category across all participants in the dataset and "Num Users" is the respective number of users that ever used an app from the respective category during data collection. The app categories were defined according to the information on Google Play Store. In addition to the 26 categories present in the table, it is also considered the category "Unknown" which comprises information on every app that is not linked with any category. This table was adapted from [61].

C

Model Optimization

Table C.1: Parameters tested for each algorithm during Model Optimization.

Algorithm	Parameters	Values
LR	C	[0.001, 0.01, 0.1, 1, 10, 100]
LDA	solver	[svd, lsqr, eigen]
KNN	n_neighbors	[1, 3, 5, 7, 9, 11, 13, 15]
	leaf_size	[1, 2, 3, 5]
	weights	[uniform, distance]
	algorithm	[auto, ball_tree, kd_tre, brute]
RF	criterion	[gini, entropy]
	n_estimators	[5, 10, 15, 20, 25, 30]
	min_samples_leaf	[1, 2, 3]
	min_samples_split	[2, 3, 4, 5, 6, 7]
	max_features	[auto, log2]
SVM	C	[0.01, 0.1, 0.5, 1, 10, 100]
	kernel	[linear, poly, sigmoid, rbf]
	gamma	[25, 50, 75, 100, 150, auto]
	degree	[1, 2, 3, 5, 10]
BOOST	loss	[deviance, exponential]
	learning_rate	[0.01, 0.1, 0.5]
	max_depth	[3, 5, 8]
	max_features	[log2, sqrt]
	criterion	[friedman_mse, mae]
	subsample	[0.5, 0.8, 1]
	n_estimators	[30, 100, 150]

D

Experimental Results

Table D.1: 1st Experiment - Mean accuracy value and the corresponding standard deviation of each algorithm, given a certain feature selection technique for every personality trait.

Personality Trait	Feature Selection Technique	Algorithm						
		LR	LDA	KNN	RF	NB	SVM	BOOST
Emotional Stability	Univariate Selection	**0.606 (0.077)**	0.596 (0.087)	**0.607 (0.117)**	0.538 (0.058)	**0.647 (0.039)**	**0.608 (0.064)**	0.588 (0.068)
	Random Forest	0.498 (0.103)	0.508 (0.096)	0.556 (0.121)	0.540 (0.093)	0.567 (0.092)	0.548 (0.128)	0.568 (0.099)
	PCA	0.497 (0.091)	0.478 (0.098)	0.509 (0.118)	0.518 (0.105)	0.509 (0.128)	0.470 (0.120)	0.534 (0.162)
Extroversion	Univariate Selection	**0.626 (0.073)**	**0.607 (0.061)**	0.567 (0.104)	0.568 (0.084)	0.587 (0.121)	**0.636 (0.121)**	0.586 (0.125)
	Random Forest	0.567 (0.102)	0.567 (0.110)	0.560 (0.100)	0.489 (0.073)	0.587 (0.113)	0.578 (0.069)	0.479 (0.075)
	PCA	0.529 (0.072)	0.559 (0.063)	0.560 (0.056)	0.491 (0.095)	0.589 (0.097)	0.490 (0.069)	0.520 (0.029)
Openness	Univariate Selection	**0.656 (0.086)**	**0.637 (0.076)**	**0.559 (0.075)**	0.431 (0.048)	**0.560 (0.069)**	0.548 (0.105)	0.508 (0.090)
	Random Forest	0.488 (0.111)	0.498 (0.114)	0.529 (0.085)	0.520 (0.066)	0.528 (0.127)	**0.597 (0.089)**	0.555 (0.232)
	PCA	0.527 (0.141)	0.528 (0.147)	0.567 (0.066)	0.530 (0.101)	0.509 (0.132)	0.567 (0.148)	0.537 (0.109)
Conscientiousness	Univariate Selection	**0.577 (0.093)**	0.539 (0.077)	0.508 (0.121)	**0.559 (0.069)**	**{0.560 (0.074)**	0.498 (0.083)	**0.579 (0.037)**
	Random Forest	0.539 (0.134)	0.538 (0.146)	0.548 (0.045)	0.549 (0.071)	0.548 (0.188)	0.518 (0.118)	0.560 (0.089)
	PCA	0.482 (0.154)	0.463 (0.155)	0.512 (0.128)	0.462 (0.143)	0.535 (0.175)	0.471 (0.174)	0.462 (0.167)
Agreeableness	Univariate Selection	**0.656 (0.116)**	**0.675 (0.084)**	0.617 (0.070)	**0.655 (0.104)**	0.597 (0.124)	**0.676 (0.069)**	0.645 (0.107)
	Random Forest	0.569 (0.0518)	0.579 (0.061)	0.520 (0.087)	0.508 (0.105)	0.590 (0.081)	0.539 (0.059)	0.500 (0.099)
	PCA	0.570 (0.085)	0.580 (0.076)	0.588 (0.024)	0.450 (0.0997)	0.531 (0.128)	0.530 (0.048)	0.501 (0.119)

Table D.2: 2nd Experiment - Mean accuracy value and the corresponding standard deviation of each algorithm, given a certain feature selection technique for every personality trait.

Personality Trait	Feature Selection Technique	Algorithm						
		LR	LDA	KNN	RF	NB	SVM	BOOST
Emotional Stability	Univariate Selection	0.448 (0.155)	**0.459 (0.136)**	**0.460 (0.051)**	0.382 (0.085)	**0.450 (0.093)**	**0.451 (0.055)**	0.372 (0.103)
	Random Forest	0.362 (0.146)	0.314 (0.127)	0.351 (0.108)	0.324 (0.077)	0.362 (0.151)	0.323 (0.049)	0.324 (0.026)
	PCA	0.341 (0.096)	0.323 (0.083)	0.374 (0.121)	0.382 (0.030)	0.364 (0.080)	0.373 (0.075)	0.374 (0.073)
Extroversion	Univariate Selection	0.462 (0.069)	**0.480 (0.047)**	**0.480 (0.080)**	0.420 (0.110)	**0.500 (0.057**	0.421 (0.080)	0.422 (0.115)
	Random Forest	0.324 (0.094)	0.383 (0.098)	0.364 (0.168)	0.303 (0.141)	0.413 (0.057)	0.394 (0.089)	0.383 (0.140)
	PCA	0.402 (0.065)	0.412 (0.040)	0.354 (0.099)	0.323 (0.045)	0.364 (0.073)	0.353 (0.094)	0.274 (0.073)
Openness	Univariate Selection	0.452 (0.113)	**0.470 (0.106)**	**0.462 (0.106)**	0.322 (0.092)	**0.531 (0.137)**	0.440 (0.135)	0.352 (0.092)
	Random Forest	0.284 (0.057)	0.285 (0.049)	0.324 (0.117)	0.246 (0.085)	0.434 (0.122)	0.255 (0.115)	0.332 (0.100)
	PCA	0.275 (0.103)	0.245 (0.063)	0.334 (0.068)	0.313 (0.099)	0.363 (0.126)	0.323 (0.078)	0.334 (0.087)
Conscientiousness	Univariate Selection	0.439 (0.084)	**0.459 (0.091)**	0.392 (0.119)	**0.450 (0.075)**	0.422 (0.081)	0.430 (0.090)	0.412 (0.066)
	Random Forest	0.410 (0.092)	0.410 (0.120)	0.362 (0.064)	0.373 (0.154)	0.392 (0.071)	0.363 (0.080)	**0.453 (0.112)**
	PCA	0.304 (0.051)	0.314 (0.051)	0.382 (0.035)	0.341 (0.138)	0.333 (0.069)	0.362 (0.046)	0.333 (0.053)
Agreeableness	Univariate Selection	0.439 (0.074)	**0.469 (0.082)**	0.422 (0.103)	**0.461 (0.077)**	**0.499 (0.106)**	**0.470 (0.113)**	0.459 (0.130)
	Random Forest	0.382 (0.070)	0.373 (0.059)	0.382 (0.077)	0.412 (0.081)	0.353 (0.085)	0.411 (0.137)	0.420 (0.139)
	PCA	0.364 (0.131)	0.385 (0.161)	0.363 (0.058)	0.383 (0.059)	0.353 (0.051)	0.433 (0.156)	0.332 (0.096)

Table D.3: Mean accuracy values and corresponding standard deviations obtained after model optimization during the 1st Experiment

Personality Trait	Algorithm	Mean Accuracy Value	Parameters	Values
Emotional Stability	Logistic Regression (LR)	0.626 (0.132)	C	0.001
	K-Neighbors Classifier (KNN)	**0.685 (0.079)**	n_neighbors leaf_size weights algorithm	3 1 'uniform' 'auto'
	Gaussian Naive Bayes (NB)	0.647 (0.039)	-	-
	Support Vector Classifier (SVM)	0.666 (0.060)	C kernel gamma degree	0.5 'sigmoid' 100 1
Extroversion	Logistic Regression (LR)	**0.696 (0.059)**	C	0.001
	Linear Discriminant Analysis (LDA)	0.607 (0.061)	solver	'svd'
	Support Vector Classifier (SVM)	0.666 (0.124)	C kernel gamma degree	0.5 'sigmoid' 75 1
Openness	Logistic Regression (LR)	0.636 (0.089)	C	0.01
	Linear Discriminant Analysis (LDA)	0.637 (0.076)	solver	'svd'
	K-Neighbors Classifier (KNN)	**0.647 (0.112)**	n_neighbors leaf_size weights algorithm	9 1 'uniform' 'auto'
	Gaussian Naive Bayes (NB)	0.560 (0.069)	-	-
Conscientiousness	Logistic Regression (LR)	**0.618 (0.061)**	C	0.01
	Random Forest Classifier (RF)	0.598 (0.035)	criterion n_estimators min_samples_leaf min_samples_split max_features	'gini' 30 2 2 'auto'
	Gaussian Naive Bayes (NB)	0.560 (0.074)	-	-
	Gradient Boosting Classifier (BOOST)	0.589 (0.065)	loss learning_rate max_depth max_features criterion subsample n_estimators	'exponential' 0.5 5 'log2' 'mae' 1 100
Agreeableness	Logistic Regression (LR)	0.685 (0.089)	C	0.01
	Linear Discriminant Analysis (LDA)	0.675 (0.084)	solver	'svd'
	Random Forest Classifier (RF)	0.684 (0.097)	criterion n_estimators min_samples_leaf min_samples_split max_features	'gini' 25 1 3 'auto'
	Support Vector Classifier (SVM)	**0.705 (0.071)**	C kernel gamma degree	100 'poly' 'auto' 3

Table D.4: Mean accuracy values obtained after model optimization during the 2nd Experiment

Personality Trait	Algorithm	Mean Accuracy Value	Parameters	Values
Emotional Stability	Linear Discriminant Analysis (LDA)	0.469 (0.121)	solver	'eigen'
	K-Neighbors Classifier (KNN)	0.469 (0.075)	n_neighbors leaf_size weights algorithm	5 1 'distance' 'auto'
	Gaussian Naive Bayes (NB)	0.450 (0.093)	-	-
	Support Vector Classifier (SVM)	**0.529 (0.096)**	C kernel gamma degree	0.1 'liner' 25 1
Extroversion	Linear Discriminant Analysis (LDA)	**0.500 (0.057)**	solver	'eigen'
	K-Neighbors Classifier (KNN)	0.500 (0.078)	n_neighbors leaf_size weights algorithm	7 1 'distance' 'auto'
	Gaussian Naive Bayes (NB)	**0.500 (0.057)**	-	-
Openness	Linear Discriminant Analysis (LDA)	0.470 (0.106)	solver	'svd'
	K-Neighbors Classifier (KNN)	0.483 (0.131)	n_neighbors leaf_size weights algorithm	3 1 'uniform' 'auto'
	Gaussian Naive Bayes (NB)	**0.531 (0.137)**	-	-
Conscientiousness	Linear Discriminant Analysis (LDA)	0.459 (0.091)	solver	'svd'
	Random Forest Classifier (RF)	**0.561 (0.119)**	criterion n_estimators min_samples_leaf min_samples_split max_features	'gini' 15 3 2 'auto'
Agreeableness	Linear Discriminant Analysis (LDA)	0.490 (0.064)	solver	'eigen'
	Random Forest Classifier (RF)	0.489 (0.070)	criterion n_estimators min_samples_leaf min_samples_split max_features	'gini' 20 1 7 'auto'
	Gaussian Naive Bayes (NB)	**0.499 (0.106)**	-	-
	Support Vector Classifier (SVM)	0.491 (0.125)	C kernel gamma degree	0.5 'sigmoid' 50 1