Mood prediction based on smartphone usage behavior

A quantitative research into the prediction of stress and anxiety levels by analyzing smartphone usage data

Master Thesis

Data Science: Business & Governance

Roderick Korthals

Student number: 2031862

Supervisor:

Dr.Hendrickson

Second reader:

Dr. Atzmüller

TILBURG UNIVERSITY

Understanding Society

Preface

This master thesis is part of the ending phase of my master Data Science: Business and Governance. By doing this research, I have realized that mood prediction is hard and that the amount of data available can very useful. Mood can be affected by various factors, and predicting mood based solely on smartphone data is ambitious. This made my research complicated and sometimes frustrating. Still, I have learned a lot from, and overall, I have enjoyed writing this thesis. I hope the results of this research provide some insights and will be helpful for researchers studying this issue in the future.

I would like to show my gratitude to my supervisor Drew Dr. Drew Hendrickson, who provided me the dataset, which made this topic possible and shedding light on solutions for barriers. Moreover, I am very grateful to my family and friends for supporting me and believing in me.

Abstract

The main goal of this study is to examine to what extent daily stress and anxiety levels can be predicted by analyzing smartphone usage data. In the literature, it became clear that smartphone use is linked to stress and anxiety, and predictive modeling has shown the potential to utilize smartphone data to successfully predict mood. Therefore, a generic and group-personalized model has been used to perform prediction tasks 1 and 2. The first prediction task examined to what extent daily stress and anxiety levels of smartphone users can be predicted by analyzing smartphone usage data. The second prediction task examined to what extent daily stress and anxiety levels can be predicted by analyzing smartphone usage data and the context when using a smartphone. Three machine learning algorithms were applied, namely decision tree, logistic regression, a support vector machine, and random forest. Overall, the models performed poorly for predicting stress and anxiety, and the results showed that the models performed better for predicting anxiety than stress. The random forest was the only model that had a moderate performance in the generic and group-personalized model. Adding external factors improved the prediction performance of the models. Moreover, the group-personalized model did improve the prediction task. The percentage of notifications and the number of sessions were the most important features in the generic model to predict anxiety. There were no crucial features identified in the generic model to predict stress. Finally, in the group-personalized models, the daily use of other (not defined) applications was of most importance when predicting stress. Daily use of Social Media, daily use of other applications, and daily use of messaging apps were the most important features to predict anxiety, although their value was limited. Since this research showed that group-personalized models had limited value in the prediction task, further research should use personalized models to predict mood. Besides, neural networks could be used, which seem to be more suitable to the prediction task.

Contents

1. Introduction

1.1 Context

The introduction of the iPhone in 2007 marked a radical change in the mobile industry, one which would eventually have a considerable impact on our day-to-day lives (Elhai et al., 2017). We have arrived at an era in which a whole generation has never experienced life without a smartphone (Twenge, 2017). Even though smartphones have, in part, made our lives easier, they have drawbacks (Elhai et al., 2017; Kuss et al., 2018). Substantial evidence has arisen, that excessive smartphone use can pose mental health hazards. Two mental health issues often linked to excessive smartphone use are stress and anxiety. Phone use in that can impose adverse effects on our lives, such as stress and anxiety, is called problematic smartphone use (Kuss et al., 2018).

A conventional theoretical framework that can explain problematic smartphone use as related to stress and anxiety is lacking. However, various studies have found multiple correlations between smartphone use and stress or anxiety (e.g., time spent on smartphone, type of applications used, notifications, sleep disturbance etc..) (Kuss et al., 2018; Elhai et al., 2017; Pivetta et al., 2019; Vahedi et al., 2018). Therefore, analyzing smartphone usage data has a great potential to detect stress or anxiety. Thus, it is interesting to examine the potential of this type of data to predict stress or anxiety.

This thesis applies predictive modelling to predict daily stress and anxiety levels of smartphone users. The study consist of two prediction tasks and two types of models. The generic model is the first type of model that uses all the data in the dataset. The second type of model, a group personalized model takes into account individuals by using only individuals with a similar degree of mood variation. In both models, prediction task 1 and 2 were performed. The first prediction task examined to what extent daily stress and anxiety levels of smartphone users can be predicted by only analyzing smartphone usage data. The second prediction task examined to what extent daily stress and anxiety levels can be predicted by analyzing smartphone usage data and the context when using a smartphone.

In this manner, it will be evaluated to what extent stress and anxiety levels can be predicted with smartphone usage data.

The datasets on which predictive modelling is applied were retrieved from Dr. Drew Hendrickson. These datasets consist of 2 months of data about smartphone usage patterns, detailed information about the applications used, and results from a 2-month survey about the moods and activities of the smartphone users.

1.2 Motivation

The problem above is worth addressing for multiple reasons. According to the World Health Organization, stress has been classified as one of the largest mental health epidemics of the 21st century (HCA, 2019). Stress can play a role in other psychological conditions such as anxiety. Moreover, according to the Anxiety and Depression Association of America, anxiety is one of the most severe mental health issues in the United States. Approximately 18.1% of the population in the United states gets affected every year. Only 36.9 % of those suffering receives treatment (Abbas, 2019). According to the World Health Organization, stress has been classified as the health epidemic of the 21st century (HCA, 2019). Smartphones have the ability to record their smartphone usage. The application of smartphone usage data in combination with machine learning techniques, shows great promise for automatically recognizing mood swings, and severe levels of stress and anxiety, and for detecting the right time for intervention (Constantinides, 2018). This can help health officials to combat mental health illnesses such as stress and anxiety.

Moreover, previous mood prediction studies have tried to predict every user’s mood by using a one-size-fits-all-model. Such an approach is inherently limited given the great degree of individual variation in how a person’s behavior and environment may affect their mood (Jacques et al., 2017, p: 2). Group-personalized models showed the potential to take into account individuals which can improve prediction performance. Only a few studies have used these type of models to predict (Palmius et al., 2018). Therefore, this study contribute to academic research by providing more insight into the potential of group-personalized models to predict mood.

To conclude, this study is of practical relevance because it can help improve detection of stress and anxiety and can help to identify the right time for intervention. Moreover, only a few studies used group-personalized models to predict mood. Therefore, this study is of scientific relevance because it provides more insight into the potential of group-personalized models to predict mood.

1.3 Research questions

The previous two sections have shown that smartphone usage has often been linked to stress and anxiety. Moreover, predictive modelling shows a significant promise in detecting stress and anxiety symptoms. Therefore this study addresses the following problem statement.

Problem statement: *Previous studies have shown that smartphone use is linked to anxiety and stress but it’s not clear to what extent smartphone use can predict stress and anxiety levels.*

To find a solution to the problem statement, the following research questions were formulated:

1 . *To what extent can stress levels of smartphone users be predicted by analyzing their daily smartphone usage?*

2. *To what extent can anxiety levels of smartphone users be predicted by analyzing their daily smartphone usage?*

3 . Do the models perform better if the context when using a smartphone is taken into account?

4 *Do group-personalized models perform better in predicting stress and anxiety levels of smartphone users than the generic models?*

1.4 Structure

This thesis is structured as follows. Chapter 2 contains a review of previous studies and discusses the relevant theory and methods to answer question 1 and to support the models created. Chapter 3 describes the experimental setup with a clear description of the dataset, the sample and variables used, the methods used, and the evaluation criteria applied. The experimental results of the analyses performed are discussed in chapter 4. Finally chapter 5 presents the answers to the research questions that results from the analyses, discusses the limitations of the study and offers recommendations for future research.

2. Theoretical Framework

In this chapter, the previous relevant literature is explored. A review of recent literature will help us identify important factors that can improve the prediction performance of the models. Moreover, the literature can provide insights into how previous analyses have been executed.

First, problematic smartphone use is discussed, as its relationship to anxiety and stress. Next discussed are related theories that can explain why people use their smartphones, and how this can lead to problematic smartphone use. Section 2.3 examines those characteristics of a smartphone, that may facilitate problematic smartphone use. Finally, Section 2.4 explains which methods were used in previous research to predict mood based on smartphone usage data.

2.1 Problematic smartphone use

Recent technological advances have led to a significant increase in the use of mobile technologies (Altuwairiqi et al., 2019, p: 1; Kuss et al., 2018, p: 1). Smartphones and the internet have become more intimately intertwined with our lives, enabling "on-the-go" access to several functionalities such as web-browsing, social networking, shopping, banking, and gaming (Kuss et al., 2018, p:1).

Because of the rising dominance of smartphones in our daily lives, more studies have been carried out that examine the side-effects of smartphones on individuals. It is observed that phone use can become problematic if it leads to detrimental effects on attention; financial issues; problematic social or academic behaviors; or symptoms of depression, stress, or anxiety ((Mitchell & Hussain, 2018). Shin and Dey (2013, p: 336) define problematic phone use as overuse or undesirable use of a smartphone that results in negative consequences to both personal and social aspects of one’s life (e.g., stress or anxiety).

The relationship between smartphone use, stress, and anxiety has received increasing attention (Vahedi et al., 2018; Kuss et al., 2018; Elhai et al., 2017). Although stress and anxiety are sometimes used interchangeably and may show the same symptoms, they refer to different constructs. Baum (1990) defines stress as the inability to cope with external demands referred to as stressors (Vahedi et al., 2018, p: 2). Anxiety is a perception, arrived at for no reason whatsoever, that there is something wrong or deficient about oneself, which will eventually lead to feelings of rejection and judgement. This kind of anxiety is often called social anxiety and is the most common form (Kuss et al., 2018, p: 2; Lukasik et al., 2019).

A significant amount of research has been conducted, that tries to explain the relationship between smartphone use and stress or anxiety. The next two sections will discuss theories that describe why people use their smartphone, what causes problematic smartphone use, and how these behavioral patterns be linked to stress or anxiety.

2.2 Habit formation and problematic smartphone use

A theory that conceptualize problematic smartphone use is the Uses and Gratification Theory (UGT). UGT discusses that people tend to use their smartphone to satisfy a need for information, interpersonal relationship, entertainment and the need to kill time which can lead to excessive reliance of a smartphone, and cause unintended habits (Wang et al., 2015, p: 5).

Elhai et al. (2017) discuss that those unintended habits can negatively reinforce themselves into problematic smartphone use via various behavioral pathways (Elhai et al., 2017). Some type of phone user have created habits that drive them to continually divert their attention to other activities. This can reach a level at which the user becomes annoying to others. When the smartphone is removed, panic attacks, or feelings of discomfort might emerge (van Deursen et al., 2015; Oulasvirsta, 2012;). Another pathway to problematic smartphone usage arises from the fact that unintended habits of checking one's phone and observing notifications also serve to provide social reassurance from friends and relationship partners. Phone checking behaviors are also related to a fear of missing out (FOMO). This term involves the reluctance to miss important information and rewards along with the need to stay continually connected with members of one's social network continually (Elhai et al., 2017, p : 253; Beyens, Frison & Eggermont, 2016). Eventually, all these pathways are considered to be routes to psychopathological symptoms such as stress and anxiety (Vahedi & Saiphoo, 2017).

In addition, other theories claim that personality traits are linked to problematic smartphone use. Takao et al. (2009) discuss that people who are impulsive or extravert are more likely to spend excessive amounts of time on their smartphone. For instance, impulsive individuals tend to have a lack of self-control and a inability to manage smartphone use. This can evolve into anti-social behavior, disinhibition, and attentional deficits. These type of behaviors are often linked to anxiety and stress (Elhai et al., 2017).

Thus, stress and anxiety can be linked to smartphone usage, which itself can be driven by various factors, that can differ between individuals. However, smartphones also possess features that can facilitate problematic smartphone use. The next section discusses this topic.

2.3 Smartphones architectures and problematic smartphone usage

Smartphones facilitate pathways through which problematic phone usage can develop (Elhai et al., 2017). First, the ease of access, combined with widespread and socially accepted use of smartphones, makes smartphones use ubiquitous. Secondly, the increasing number of phone functions (applications), make users more reliant on the technology and incentivize smartphone usage over other options such as analog devices. Third, apps are designed to make users prolong their usage or come back to the app (e.g., infinitive scrolling and, notifications (Noe et al., 2019, p: 61). All of these aspects can facilitate habitual behavior (van Deursen, 2015). The previous section demonstrated that such habits can lead to problematic smartphone use, which has detrimental effects on users' mental health (e.g., anxiety or stress).

Research has shown that some application types are more linked to stress and anxiety than others. For instance, social media and gaming applications feature characteristics that can enhance habitual checking behavior and seeking reassurance and can provoke FOMO. Noe et al. (2019), found that FOMO especially applies to all social media apps and is less pronounced with other types of applications (e.g., utility, lifestyle). Moreover, Montag et al. (2019) find that gaming and social media apps in particular are designed in a manner that can cause problematic smartphone use.

In this section, it has become clear that smartphones include features that can trigger problematic smartphone usage and that some type of features are more linked to stress and anxiety than others.

2.4 Related work

Various studies have been conducted, which uses smartphone usage data to predict mood. For instance, Ferdous, Osmani, and Mayora (2018) tried to predict stress levels of people at the workplace with smartphone usage pattern data. They used a Support Vector Machine algorithm. Becker et al., (2016) predicted daily mood levels by analyzing smartphone usage patterns (e.g., app categories, screen time). They also used features that take into account the context in which the smartphone was being used (e.g., type of activity, type of social environment, or type of day). Their study showed that adding contextual data to the prediction task significantly improved the prediction performance. In their analyses, they used Bayesian Hierarchical networking analyses and a Support Vector Machine. Moreover, Jacques et al. (2017) and Pratap et al. (2017) used neural networks and a Random Forest algorithm to predict mood levels with smartphone data. They also used context features (e.g., GPS data, activity) in their models.

These studies have in common that they used generic and personalized modes to predict mood levels (Jacques et al., 2017; Becker et al., 2016; Ferdous, Osmani & Mayora, 2018; Pratap et al., 2017). Personalized models are models that take into account individuals when predicting mood. Jacques et al. (2017, p: 2), claimed that one of the shortcomings of traditional mood prediction systems is that it attempts to predict every's person's mood using the same, one-size-fits-all model (population norm). Such an approach is inherently limited due to a high degree of individual variation in how a person's behavior and environment may affect their mood. For example, not every user reacts in the same way to excessive social media use. In these studies, the personalized models have shown greater predictive performance than the generic model. For instance, in the study of Ferdous, Osmani & Mayora (2018), the personalized model had an average accuracy performance of 75 %, whereas the generic model had an accuracy performance of 54 %.

A drawback of those models is that they require an abundance of longitudinal data (Jacques et al., 2017). Palmius et al., (2018) tried to tackle this problem by creating group-personalized models that only incorporate certain groups of people who show similar mood characteristics. The study showed that group-personalized models performed better than the generic model.

Previous research has shown that various algorithms and methods were used to predict mood with smartphone usage data in which personalized models had the highest prediction performance. Nonetheless, those models require an abundance of longitudinal data, which is not the case in this thesis. For this reason, a group personalized model is used in this thesis. At last, previous research showed that prediction significantly improved by adding contextual data to the prediction task. Therefore, these features will be used in prediction task 2.

3. Methods

3.1 Raw dataset

This study uses multiple datasets, which were retrieved from Dr. Drew Hendrickson. Those three datasets represents phone use data, data about the characteristics of used apps, and mood data representing results of an online survey. These datasets were retrieved from Dr. Drew Hendrickson. An overview of the datasets used and relevant information about them can be found in Table 1.

Dataset	**Time period**	**Sample size**	**Number of features extracted**
Phone use dataset	2019-02-21 – 2019-03-19	586.792	28 features
Meta-data about applications	Not applicable	1.748	18 features
Survey data about applications	2019-02-21 – 2019-03-19	9.612	13 features

Table 1: Raw dataset and relevant information

The phone use data represents 2 months information about the smartphone use of 124 users. The information includes the type of application used, session number, the start time of a used app, the end time of a used app, and the battery level. The application dataset consists of detailed information about the application used in the phone use dataset. Finally, the mood dataset consists of data collected from an online survey. The 124 phone users from the first dataset participated in this survey. The participants were required to complete four times per day. Eventually, the dataset consists of Likert-scale scores from 0 to 5 indicating the respondents mood, type of activities done during a day, and the social context of the day (e.g., family, friend, public, private).

3.2 Sample

3.2.1 Sample for the generic models

This thesis makes use of both a generic and a group-personalized model to predict daily stress and anxiety levels of smartphone users. In the generic model, the entire population of the Feature table have been used. The sample for the generic dataset includes aggregated daily features (features computed over a 24-hour period) about smartphone users' mood level, and features describing their smartphone usage (e.g., amount of time spent on application, amount of applications used). The two target variables are the aggregated daily stress and anxiety scores, which are categorical variables representing anxious or not anxious, and either stressed or not stressed. These scores were extracted from the survey data. Eventually, the dataset consisted of 36 features and 2275 observations.

3.2.2 Sample for group-personalized models

In the literature, it became clear that similar studies make use of personalized models that seem to improve prediction. These models are called “personalized” model because individuals have been taken into account in creating the model. However, these approaches assume that an abundance of individual longitudinal data is available (Pratap et al., 2017; Ferdous, Osmani & Mayora, 2018). That is not the case in this study.

A compromise between population-level models (generic) and fully personalized models is a group-personalized model. In a group personalized model, a model is created that consists of individuals with similar mood characteristics (Palmius et al., (2018). Only smartphone users who show a similar degree of variations in their mood are included in the prediction task. In a explanatory analysis, it became clear that some smartphone user’s showed a higher degree of variance in stress and anxiety levels. On the other side, some smartphone users did not show any variation in their mood. It is assumed that by leaving those users out of the model, noise will be reduced. Hopefully, when the models are provided with data only on users who show a similar degree of mood variability, models will more easily recognize patterns (Palmius et al, 2018). The group-personalized model, included data from only those smartphone users who showed an average mood variation above 0.5 and had at least 10 days of longitudinal data. Eventually, the sample of the group-personalized model to predict stress consists of 1381 observations and 36 features. The group-personalized model to predict anxiety consists of 779 observations, 36 features. Table 2 provides information about all the features used.

Nr.	Variables	Retrieved from which dataset	Description	Number of variables
1	Used applications **(**prediction task 1,2)	Phone use and application datasets	Numerical variables reflecting number of apps used per category per day	9
2	Time spent on application (prediction task 1,2)	Phone use and application	Numerical variables reflecting minutes spent on application per category per day	9
3	Notifications (prediction task 1,2)	Phone use dataset	Numerical variable reflecting the percentage of the entire phone activity during a day, in which a notification has been received	1
4	Number of sessions (prediction task 1,2)	Phone use dataset	Numerical variables reflecting number of sessions per day	1
5	Weekday/weekend (prediction task 2)	Mood dataset	Dichotomous variable: Weekday (0), Weekend (1).	1
6	Activities and social context (prediction task 2)	Mood dataset	Numerical variable: Detailed information in section 3.3.2	11

Table 2: Input variables used in analysis and short description

3.3 Preprocessing

Little cleaning was required for the phone use dataset and application dataset. On the other hand, there were some missing and extreme values in the mood data. The missing values were caused by respondents who had started a survey but had not managed to complete it. These missing values were ultimately deleted from the dataset because these rows did not possess any anxiety and stress scores. After the cleaning process, data from every user was aggregated in daily usage patterns and their reported mood. In the phone use datasets, new features were created to indicate the number of application categories used and time spent on an application category. Other categories were also created that show the number of notifications sent and the current activity when completing the survey.

3.3.1 Target variable

Stress and anxiety

In this thesis, the features to be predicted are whether a smartphone user is stressed or not, and whether a smartphone user is anxious or not. In the survey, they could indicate their stress and anxiety levels four times per day. The scores for their stress and anxiety levels were recorded on a Likert scale ranging from 0 (not stressed or anxious at all) to 5 (very stressed or anxious). These Likert scores were eventually aggregated into average daily stress and anxiety scores. Moreover, the decision was made to divide these scores into a binary target variable in which 0-2 represents not stressed, and 3-5 represents stressed. These scores were all reported in the period from 21 February 2019 to 19 March 2019.

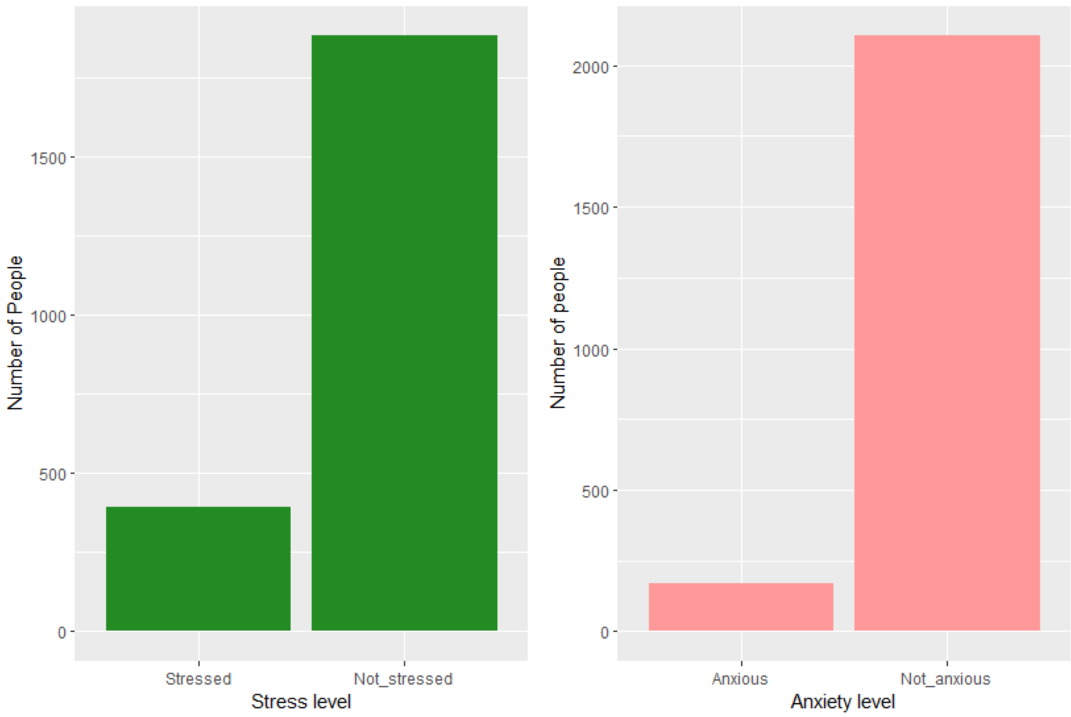

Figure 1: Distribution of target variable

As can be seen in Figure 1, there were many more, not-anxious, and not-stressed smartphone users than anxious and stressed users. This issue needed to be considered when analysis was performed. Section 3.4.2 elaborates on this topic.

3.3.2 Input variables

Smartphone usage

This study has a particular interest in the type of application and its effect on stress and anxiety levels. To create meaningful features, the decision was made to create nine application categories: Utility (camera, calculator), Productivity (Gmail, Grip), Gaming and Entertainment (Angry Birds, Netflix) , Social Media (Snapchat, Instagram, Facebook), Messaging (Whatsapp), News and Sports (Nu.nl, Voetbalzone), Browser (Internet browser), Lifestyle (Tinder, Zalando) and Other (not defined).

Moreover, notifications and the number of sessions a day were also used as features that described a user's smartphone usage. The notification feature represented the percentage of the entire phone activity (used applications) during a day, in which a notification was received. It is clear from the literature that notifications can trigger smartphone usage, which eventually can cause increased stress levels. Furthermore, the number of sessions is a good indication of how often a user turned the smartphone one and off during a day. This can indicate the stressed and anxious behavior of a smartphone user.

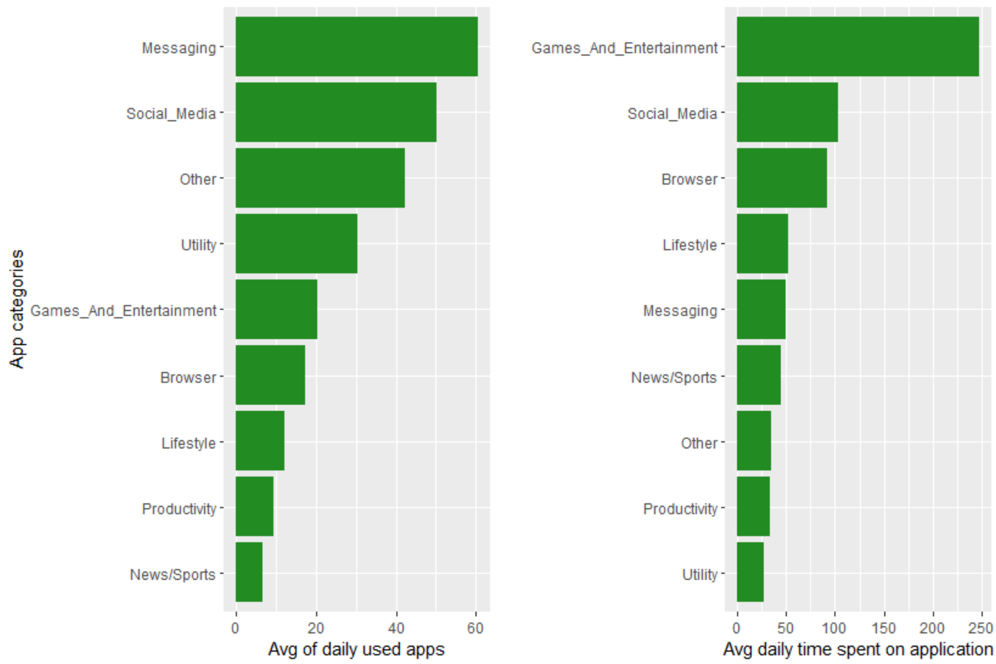

Figure 2: The average amount of time spent and the average number of app categories used

Activities, social context and type of day are three categories that were added to the models in the second prediction task. In this prediction task, it will be examined if adding external factors to the model can improve prediction. The literature makes clear that people have different personality trait and react differently to the same stimuli. Therefore, it was assumed that the type of activity (study, sport, social, class, etc.), the social context (with friend, family, private etc.) and the kind of day (weekday or weekend) may influence the mood of the smartphone user.

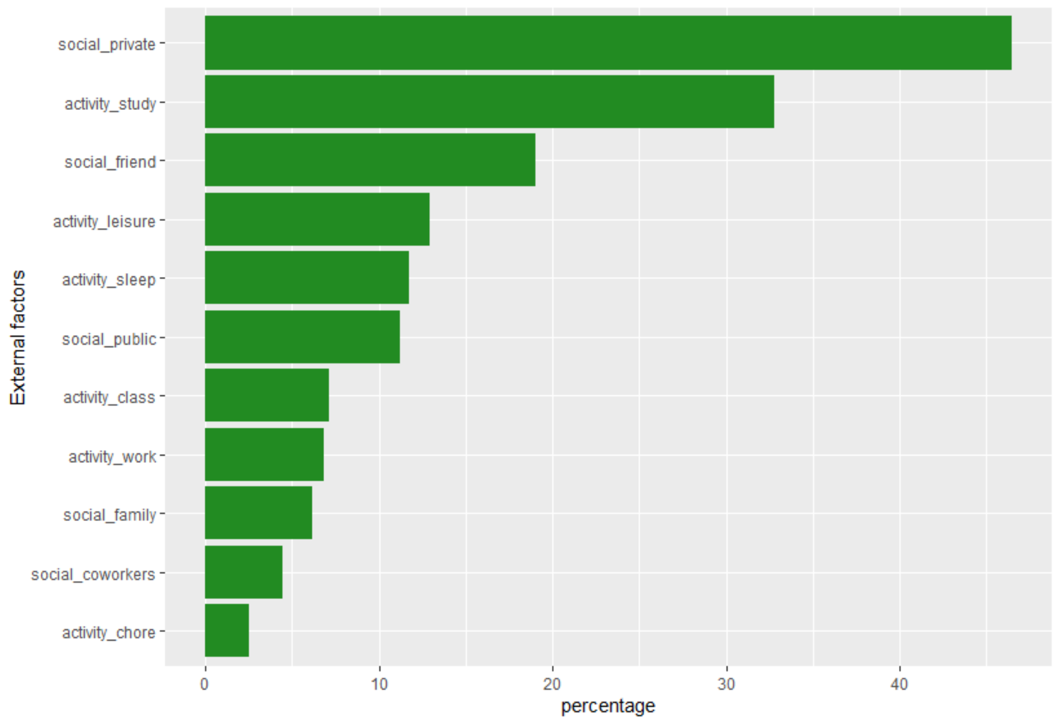

Figure 3: Descriptive statistics external factors

3.4 Experimental set up

3.4.1 Training and test set

Before a model can be trained, the dataset must be split into a test and a training sets. The dataset was split, with 70% training set and a 30 % in a test set. The training set was used to train the model, and the test set was then used to see how well the model performed on new data.

3.4.2 Unbalanced dataset

As can be seen in Figure 3 and 4, the target variables suffer from severe class imbalances. The problem of imbalanced data is that the minority class is more often misclassified in comparison to the majority class. The problem increases when this minority class contains information that is essential (Tantithamthavorn, Hassan & Matsumoto, 2018). Various solutions have been proposed to deal with this problem. Among the mainstream solutions are over-sampling and under-sampling. In the under-sampling process, data for the majority class are reduced or eliminated to balance the class distribution. By contrast, over-sampling is done by adding data to the minority class (Sonak & Patankar, 2015, p: 339).

These methods seem to cause some derived problems. Under-sampling may discard some useful examples for modeling the classifier. Mainly when the ratio of imbalance is high, then more cases need to be removed, which leads to a reduction of the majority class. This may affect the generalization ability of the classifier (Fernandez et al., 2018, p: 864).. Alternatively, an over-sampling method has frequently been used as a solution. This method tackles the class imbalance by overrepresenting the minority class examples. The advantage of this method is that there is no loss of data. The disadvantage is that it may lead to overfitting (Fernandez et al., 2018; Sonak & Patankar, 2015, p: 339).

In this thesis, the Random-Sampling examples algorithm (ROSE) is used, which is an alternative to standard random over-sampling (Fernandez et al., 2018). This technique uses a smoothed bootstrapping approach to draw artificial samples from the feature space neighborhood and minority class. ROSE combines over-sampling and under-sampling by generating an augmented example of the data. The ROSE technique features three steps. First, the minority class is resampled using a bootstrap resampling technique to remove modules of the majority class. Second, the minority class is also resampled using a bootstrap technique. Third, new synthetic data is created in the neighborhood of the feature space (Tantithamthavorn, Hassan & Matsumoto, 2018, p: 4).

Multiple studies have been conducted which have studied the impact of those rebalancing techniques. For instance, Ahn and Ahn (2018) found evidence that ROSE improves the performance in predicting a binary classifier (bankrupt or not). Thus, the ROSE algorithm was used in this study to increase the performance of the models.

3.4.3 Algorithms

In this thesis, the prediction task was to assess if whether someone is stressed or anxious or not stressed or anxious. To do this, the target variables, which were initially a score between 0-5, were transformed into binary classifier of not-stressed/stressed, and not-anxious/anxious. The respondents were classified using the algorithms Decision Tree, Logistic Regression, Support Vector Machine, the Random Forest classifier. The following sections elaborate on these algorithms.

Decision tree

The decision tree is a frequently used classifier for nominal, binary, and numerical target values. A decision tree, also known as a classification tree, is a tree-like model in which a complex decision-making process is divided into a collection of several more straightforward steps. The goal is to create a classification model that predicts the target variable of the target attribute based on several input variables. Moreover, growing decision tree's beyond a certain level can result in overfitting. This could decrease the performances of the model. Finding the right value for the complexity parameter can overcome overfitting (Patel, 2015).

An advantage of the decision tree to other approaches is that it is meaningful and easy to interpret. Therefore, the decision tree was selected as the baseline algorithm because baseline models have the purpose of explaining the data in a simple manner. A disadvantage of the decision tree, however, is that it performs weakly when the complexity (large number of features) of the model is great. Therefore other algorithms were also used that can better address model complexity (Patel, 2015).

Logistic regression

Ordinary linear regression was not used because the target variable is not continuous. Logistic regression (LR), by contrast, is suitable for problems with discrete outcome variables. LR applies maximum likelihood estimation after transforming the dependent variable into a logit variable (the natural log of the odds of the dependent variable occurring or not). In this way, LR estimates the probability of a particular event occurring (Liu et al., 2011, p: 2).

LR can be used to overcome overfitting through the use of L2 regularization. L2 regularization penalizes weight variance and makes a trade-off between a fitted model and a simple model. Therefore, this model can also support the feature selection process and helps select the best performing features(Liu, Chen & Ye, 2009). A disadvantage of the LR is that it’s unable to solve non-linear problems.

Support vector machine

The support vector machine (SVM) is a machine learning method that is widely used for linear and non-linear problems with binary outputs. In the SVM, the original input space is mapped into a high dimensional feature space, and in these feature space, the optimal hyperplane is determined. The optimal separating hyperplane is the one that correctly classifies all the data while being farthest away from the data points. The optimal hyperplane is claimed to maximize the generalization of the model. An advantage of the SVM is that it successfully solves the problem of high dimensionality (Inoue & Abe, 2001).

Random Forest

The random forest algorithm is an extension of the decision tree algorithm that uses several prediction trees that are less tolerant to noise and uses a random selection of features in splitting the trees. A random forest is a voting procedure for the most popular class among a large number of trees. Thus, a random forest is composed of a set of decision trees. In the random forest algorithm, a prediction is made by taking an average of the answers. The advantage of the random forest is that it can handle a very large number of input variables without overfitting (Ghatasheh, 2014, p: 21: Biau, 2012, p: 1).

In choosing the optimal number of trees, a trade-off must be made between performance and processing time. A small number of trees can decrease the performance of the model, whereas a large number of trees can increase computational cost. Perez and Baranauskas (2012, p: 166) studied the optimal amount of trees to use in a random forest. They concluded that the optimal number ranges between 64 and 128 and that growing more trees does not improve results but does increase computational costs. Therefore, a number of trees were tested, ranged between 64 and 128, to optimize the parameters of the random forest.

3.4.4 Evaluation criteria

Similar studies have used multiple evaluation criteria to assess the performance of their models. A study by Constantinides et al. (2018) used accuracy as a measure to evaluate the performance of their models at predicting whether or not people were depressed. Moreover, Becker et al. (2016) used root mean squared error (RMSE) to assess the mood level of smartphone users, whereas Pratap et al. (2017) used the area under the curve (AUC) score as a model evaluation.

In this study, the main goal is to correctly predict whether or not people are stressed or not. Miss-classifying on either sides can be costly, because misclassifying an anxious person as not anxious can cause no treatment when treatment is needed. On the contrary, classifying a not anxious person as anxious can cause unnecessary medical prescriptions, which can increase the medical costs of the smartphone user (Ling, Sheng & Yang, 2006). Therefore, it is necessary that the models have high performance on both sides.

Accuracy is a commonly used performance measure for machine learning although it is not suitable for imbalanced datasets. In such cases, one class tends to dominate the data, and a classification model could classify all the data with the majority class, reaching a reasonable level of accuracy even though such results would be misleading since the examples in the minority class would be ignored entirely (López-Santamaría et al., 2019, p :4). Kappa is a better measurement when the dataset is imbalanced, and misclassifying on both sides is costly. Kappa is a metric that compares observed accuracy and expected accuracy (which is equal to random chance). It takes a value between 1 (best) and 0 (worst). By taking random chance into account, the metric becomes less misleading (Cohen, 1968). Since this study deals with an unbalanced dataset, Kappa was used as an evaluation metric.

Another performance metric that is a relevant measure in our study is AUC. The AUC score is derived from the Receiver Operating Curve (ROC), which is generated by changing a set of trade-off points between true positives and false positives. Thus, the estimated AUC is statistically interpreted as the probability that the classification model will correctly classify people (Miao, Miao, Miao, 2015). The AUC index ranges from 0.5 to 1, in which 0.5 indicates a random and weak model, and 1 represents a perfect classification performance. Similar to the Kappa score, the AUC score is a better evaluation criterion when the dataset is imbalanced, because it does not put more emphasis on one class over the other (Ramyachitra & Manikdan, 2014). Therefore, the ROC evaluation method was also used in this thesis.

3.5 Implementation

The merging, transformation, and pre-processing to produce the final dataset were completed using Rstudio (version 3.3.6). The models and plots were created in Rstudio with the CARET package and ggplot2.

4. Results

The goal of this study is predict daily stress and anxiety levels of smartphone users with smartphone usage data. A generic and group personalized model are used that both perform prediction task 1 and 2. Section 4.1 discuss the application of the ROSE algorithm and the parameters used. Section 4.2 describes the results of the generic and group personalized model performing prediction task 1 and 2. Section 4.3 discuss which features are most important in predicting anxiety and stress levels, according to the best performing algorithm. Section 4.4 then summarizes the results.

4.1 Balancing the dataset

The first task was to create two models that perform a binary classification of stressed and not stressed smartphone users, and anxious and not anxious smartphone users. By applying the ROSE technique, the unbalanced dataset was turned into a balanced dataset. Through an exploratory analysis, these models were shown to perform better with a balanced training set. Therefore, it is decided to apply the ROSE algorithm to all the used models. In the balanced training set to predict anxiety there were 764 anxious smartphone users and 829 not-anxious users. The balanced training set to predict stress had the same division. Figure 5 shows an example of the distribution of anxious and not anxious smartphone users before sampling and after sampling.

Figure 4: The results of the ROSE algorithm on the training set

4.2 Parameter optimization

Feature selection techniques were carried out to reduce the dimensionality of the feature table. Reducing the number of features can reduce overfitting and improve the performance of the models. Various methods of achieving this have been tried, such as a principal component analysis, evaluating correlations between variables, L2 regularization, and assessing the importance of the features of the baseline algorithm. In the appendix, elaborated information can be found about the results of the performed feature selection techniques. Unfortunately, all these methods failed to identify the most prominent features. Therefore, all features were retained in the models. Moreover, a grid search of 10 fold cross validation were applied in used algorithms. The decision tree algorithm was also pruned to reduce complexity. At last, the number of trees tested for the random forest algorithm ranged between 64 and 128 trees. The highest performance of random forest was achieved with 128 trees.

After optimizing the parameters, the Kappa and AUC scores were retrieved for the models. Eventually, two models were used and two prediction tasks were performed to predict daily stress and anxiety levels of smartphone users.

4.3 Performance models

In this section, the results of prediction task 1 and 2 will be discussed. Note that the group-personalized models use features from only those users with fluctuations in their moods.

4.3.1 Performance of the generic models

Table 3 lists the scores the generic models predicting stress and anxiety with the AUC and Kappa score of each algorithm.

	Stress				**Anxiety**			
	Prediction task 1		**Prediction Task 2**		**Prediction task 1**		**Prediction task 2**	
Algorithms	**Kappa**	**AUC**	**Kappa**	**AUC**	**Kappa**	**AUC**	**Kappa**	**AUC**
Decision Tree	0.019	0.55	0.09	0.59	0.13	0.63	0.13	0.63
LG	0.1	0.55	0.07	0.59	0.02	0.56	0.04	0.55
SVM	0.16	0.64	0.19	0.68	0.18	0.72	0.25	0.73
Random Forest	0.23	0.69	0.27	0.73	0.21	0.75	0.26	0.77

Table 3: Results of the generic model

Stress

As can be seen in Table, none of the algorithms performed well. The baseline algorithm (Decision Tree) (AUC = 0.55, Kappa = 0.019) and the LR (AUC = 0.1, Kappa = 0.55) had the lowest performance. The random forest was the best performing algorithm, with an AUC score of 0.69 and a Kappa of 0.21. Moreover, adding external factors to the generic model improved prediction. Overall, the random forest (AUC = 0.73, Kappa = 0.27) in prediction task 2 is the only model that performs moderately with an AUC score above 0.7,

Anxiety

It can be derived from table 1 that the logistic regression and the decision tree (AUC = 0.63, Kappa = 0.13) had the lowest performance. The SVM (AUC = 0.72, Kappa = 0.18) and the Random Forest (AUC = 0.75, Kappa = 0.21) had a significantly higher score than the baseline algorithm. Adding external factors to the generic model improves the prediction of the Random Forest (AUC = 0.75, Kappa = 0.21) and SVM (AUC = 0.77, Kappa = 0.26).

It becomes clear that the non-linear models (random forest and SVM) performed better than the linear model (LR), which indicates that the relationship between the target variable and features is probably non-linear. Moreover, a reason that the generic models predicted anxiety better than stress is that anxiety is a more of a steady state mood. In contrast, stress has more day to day fluctuations. Therefore it is probably easier for the model to predict anxiety.

4.3.2 Performance of the group-personalized models

Table 4 lists the scores of the group-personalized models predicting stress and anxiety with the AUC and Kappa score of each algorithm.

	Stress				**Anxiety**			
	Prediction task 1		**Prediction Task 2**		**Prediction task 1**		**Prediction task 2**	
Algorithms	**Kappa**	**AUC**	**Kappa**	**AUC**	**Kappa**	**AUC**	**Kappa**	**AUC**
Decision Tree	0.07	0.52	0.1	0.55	0.14	0.64	0.2	0.65
Logistic Regression	0.13	0.59	0.07	0.59	0.1	0.58	0.14	0.63
SVM	0.11	0.61	0.18	0.68	0.22	0.61	0.22	0.7
random forest	0.19	0.65	0.27	0.73	0.25	0.7	0.25	0.74

Table 4: Results of the group-personalized model

Stress

As can be seen in Table, none of the algorithms performed well. The baseline method (AUC = 0.52, Kappa = 0.07) and the LR (AUC = 0.59, Kappa = 0.13) had the lowest scores. Adding external factors to the model, improved the prediction performance of the random forest and SVM. Overall, the random forest (AUC = 0.73, Kappa = 0.27) in prediction task 2 was the only model with an AUC above 0.7 and a Kappa above 0.2 which indicates that it is the only fair model to predict stress.

Anxiety

Table 10 indicates that the group-personalized anxiety model performs poorly in predicting anxiety. The decision tree (AUC = 0.64, Kappa = 0.14), SVM (AUC = 0.61, Kappa = 0.22), and the LR (AUC = 0.58, Kappa = 0.1) had the lowest scores in prediction task 1. The random forest is the only model that performs moderately with an AUC score of 0.7 and a Kappa of 0.25. Adding external factors, only had a large effect on the SVM (AUC = 0.7, Kappa = 0.22). Overall , the random forest algorithm in prediction task 2 (AUC = 0.74, Kappa = 0.25) is the best performing model to predict anxiety.

In comparison, the group personalized model did not show any improvements over the generic models. The random forest algorithm even performed worse than the generic model in predicting stress and anxiety.

4.4 Feature importance

This section evaluates the importance of the features in the generic and personalized models. The feature importance will be established by looking at the Gini index of the Random Forest, since this was the best performing model in the generic and group personalized model. Moreover, a correlation analysis was carried out between the target variables and the features. However, no strong correlations (>0.5) were found. The correlations are found in the appendix. The next two sections will discuss the most important variables for the generic and group personalized models in predicting stress and anxiety levels of smartphone users.

4.3.1 Generic Model

Figure 5 shows the feature importance of the generic model predicting stress and anxiety by using the Mean Decreased Gini (MDG). A larger MDG value indicates high purity of the feature, which means that the features are more important to the classification task (Li et al., 2019, p : 5725)

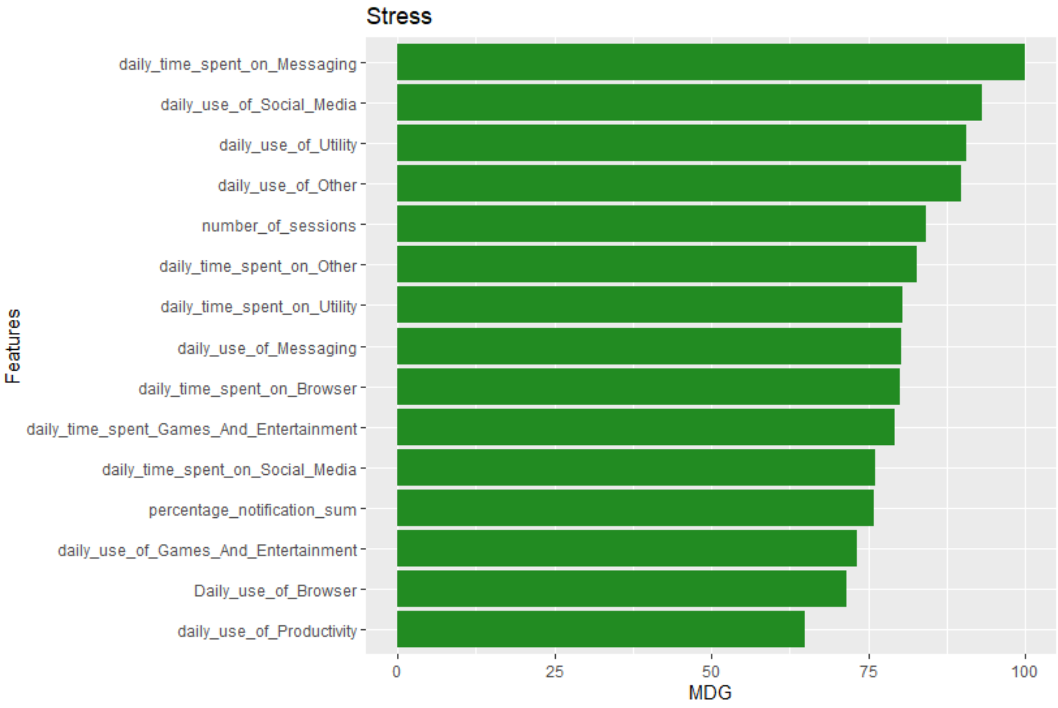

Figure 5: MDG index of generic model predicting stress

Stress

Figure suggests that the daily time spent on messaging, the number of sessions, and daily time spent on Social Media, and daily time spent on other (not defined applications) are the best predictors of stress. However, it can be seen that no wide gaps exist between the bars. This indicates that most of the features have equal importance. This could also explain the reason why the feature selection process did not provide any useful results. At last, the importance of the external factors is limited even though the external factors increased the performance of the generic model,

Anxiety

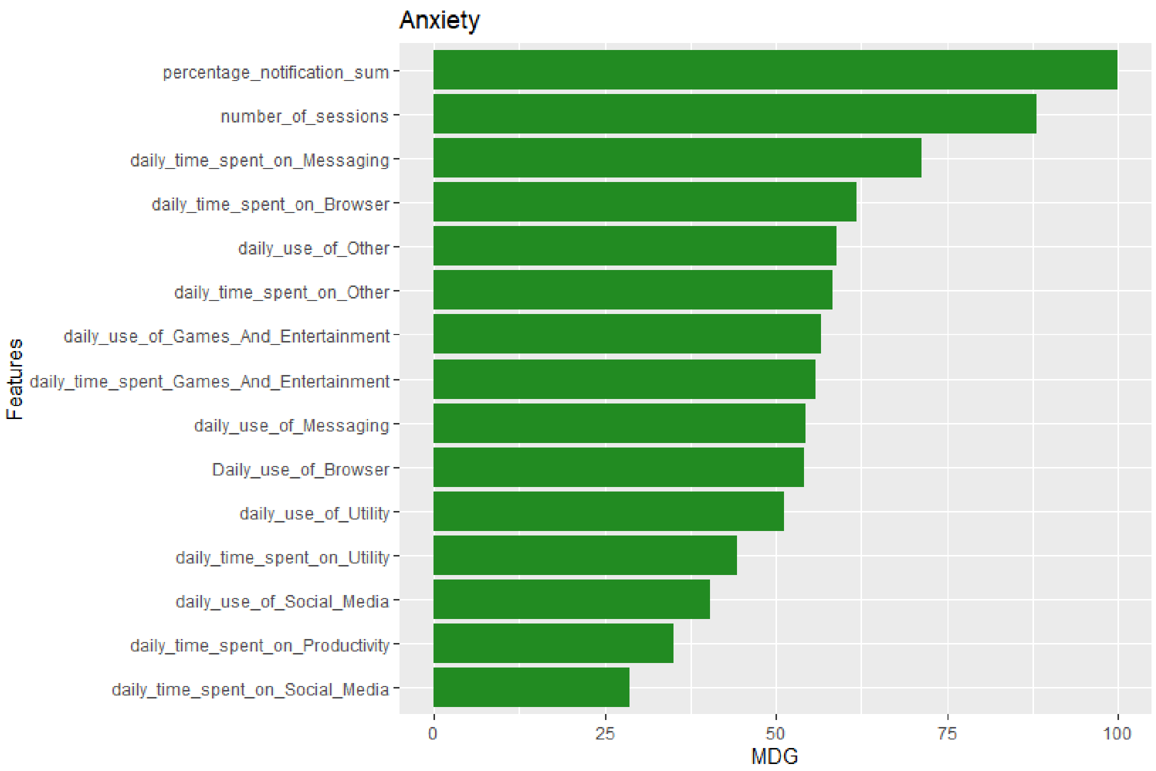

Figure 6: MDG of the generic model predicting anxiety

As can be seen in figure 6, the most important features to predict anxiety are the percentage of notifications and the number of sessions. The other features are considered to be less important. Although the external factors increased the performance of the generic anxiety model, those features were not considered to be important by the gini index.

4.3.2 Group-personalized model

Stress

Figure 7 illustrate the 15 most important features to predict stress in the group-personalized model.

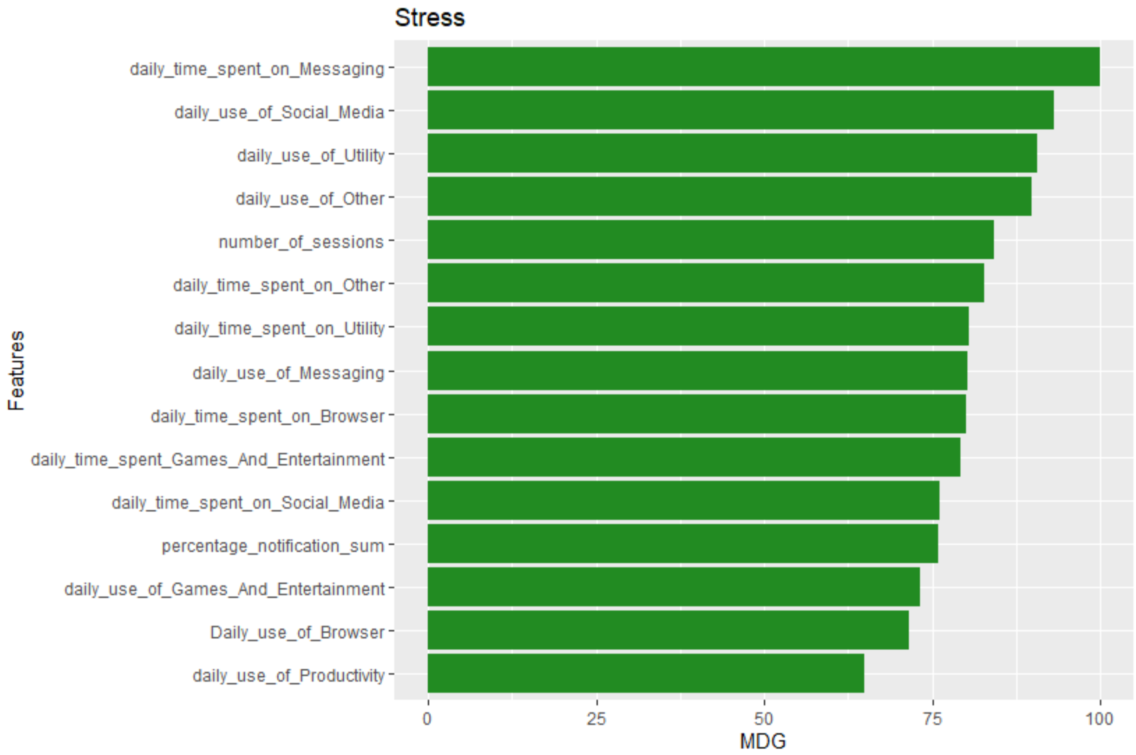

Figure 7: MDG of the group-personalized model predicting stress

In figure 7, it becomes clear that the feature daily use of other applications is the most important feature to predict stress. This feature increases in importance in the group personalized model. Other features in figure 7 show no distinctive importance. This has also been the case in the generic model. At last, the external factors show limited importance in the group personalized model

Anxiety

Figure 8 shows the MDG of the features in the group-personalized model for predicting anxiety.

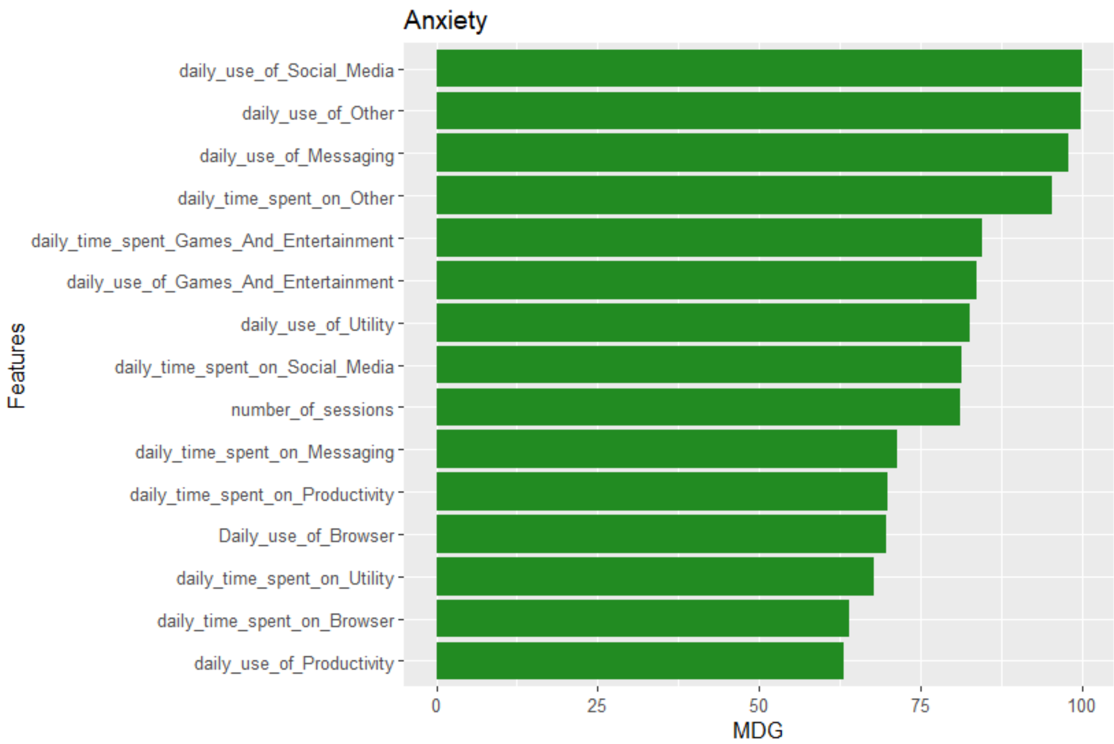

Figure 8: MDG of the group-personalized model predicting anxiety

In the group personalized model, it seems to be clear that the daily use of social media, daily use of not defined applications, daily use of messaging are the most important features to predict anxiety. However, in this graph it becomes evident that the differences between features is very limited. In predicting anxiety, the external factors also show limited importance.

In overall, the group personalized model seem to change the importance of features slightly. In the group personalized model to predict stress, the importance of other features becomes evident. It can be concluded that most of the features are unrelated to the target variables. This made the prediction process harder. The next figure will illustrate this:

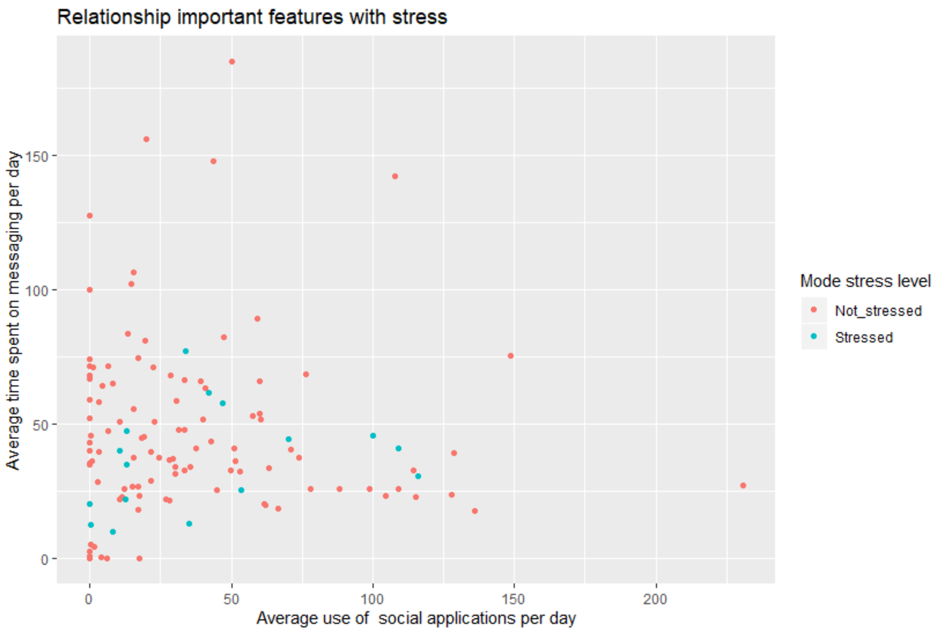

Figure 9: Relationship between two most important features in the generic model with the target variable

Figure 9 illustrates the relationship between the two most important features and the target variable. The red dots reflect people who are more frequently not stressed , and the blue dots reflect people who are more frequently stressed. A cluster of dots would suggest a relationship between the two most important features and the target variable. Nonetheless, a cluster is missing, which means that no relationship exists between the two most important features and the target variable. This can explain the poor performance of the model.

4.4 Summary results

Considering the generic models, it can be concluded that the random forest algorithm had the most predictive power by adding external factors. The random forest performed better in predicting anxiety (AUC = 0.77, Kappa = 0.26) than stress (AUC = 0.73, Kappa = 0.27). A reason can be that anxiety is a more of a steady state mood. Therefore, the models would probably be more accurate at predicting anxiety

Considering the group-personalized models, the models performed worse than the generic models. An increase in dimensionality is probably a reason that the models performed worse than the generic models. Thus, considering only people who showed notable variance in their moods, did not seem to improve predictions.

The feature importance figures showed that the influence of the variables is limited. Only notifications and the number of sessions were identified as important in the generic model to predict anxiety. The figures in section 4.4.2 indicate that a clear relationship between the features and the target variable is missing.

5. Conclusion

This Chapter discusses the research conducted and the results. First, the research questions that were formulated in section 1.4 are answered. The contribution of this research is discussed in section 5.2. The limitations of this research are discussed in section 5.3, as are recommendations for further research. The chapter concludes in section 5.4.

5.1 Answers to the research questions

This study addresses the following problem statement:

Problem statement: *Previous studies have shown that smartphone use is linked to anxiety and stress. However, it's not clear to what extent stress and anxiety levels can be predicted by analyzing smartphone usage data.*

To find an answer to the problem statement, four research questions were formulated.

1. *To what extent can daily stress levels of smartphone users be predicted by analyzing their daily smartphone usage?*

2. *To what extent can daily anxiety levels of smartphone users be predicted by analyzing their daily smartphone usage ?*

3. *Do the models perform better if the context when using a smartphone is taken into account?*

4. *Do group-personalized models perform better in predicting daily stress and anxiety levels of smartphone users than the generic models?*

The remainder of this section discusses the answers to these research questions, which collectively answer the problem statement.

Research question 1: To what extent can daily stress levels of smartphone users be predicted by analyzing their daily smartphone usage?

The first research question evolves around the performance of prediction task 1 to predict daily stress levels of smartphone users. The outcomes of prediction task 1 indicate that the SVM and random forest outperform the decision tree and LR. This indicates that the prediction problem is non-linear. Overall, none of the algorithms have an AUC above 0.7. This indicates that all the algorithms performed poorly in predicting stress levels. Considering other studies, it is not surprising that the generic models performed poorly. In similar studies that also used generic models, no models performed with an accuracy above 0.7 (Constantinides et al., 2018; Jacques et al., 2017; Pratap et al., 2017; Becker et al., 2016). It can be concluded that daily stress levels cannot be fairly predicted by analyzing their daily smartphone usage.

Research question 2: To what extent can daily anxiety levels of smartphone users be predicted by analyzing their daily smartphone usage?

The generic model to predict anxiety performed better than the generic model predicting stress. A reason may be that stress has more day to day fluctuations, whereas anxiety is a more of a steady state mood. Therefore, the models would probably be more accurate at predicting anxiety. The random forest and SVM performed the best with AUC scores above 0.7, indicating that these models perform moderately. Thus, it can be stated that daily anxiety levels can moderately be predicted with a random forest or a SVM.

Research question 3: Do the models perform better when external factors are taken into account?

In prediction task 2, external factors are added to the model. Considering the generic model, it became clear that adding external features to the prediction task improved the performance of the random forest and SVM in predicting stress and anxiety. External factors had less impact on the performances of the decision tree and LR. Moreover, external factors had more of an impact on the group personalized model. The SVM significantly performed better, and the random forest also noticed an increase in performance. The influence of external factors on the decision tree and LR is limited. In conclusion, the external factors improved the prediction performance of the generic and group-personalized model in predicting stress and anxiety.

Research question 4: Do group-personalized models perform better in predicting stress and anxiety levels of smartphone users than the generic models?

To answer question 4, a group-personalized model was created, that selected only smartphone users with the same high mood variability. Smartphone users who do not seem to exhibit a high degree of variability in their mood and did not provide longitudinal data of more than 10 days were excluded. It was assumed that by limiting the data in this manner, the model could more accurately detect patterns that the generic model would miss. Nonetheless, the group personalized model did not improve performance of the model and decreased the performance of the random forest and SVM. This may have been because of the increase in dimensionality.

5.2 Contributions

A small number of studies have used group personalized models to predict mood (Palmius et al., 2018). This study provided more insights into the usability of using a group personalized model to predict daily stress and anxiety levels of smartphone users.

5.3 Limitations and further research

The intention of this study was to create a generic and personalized model and compare their predictive performances. However, a limitation of this study was a lack of individual data to create a fully personalized model. Therefore, another approach has been adopted, which showed limited value. A recommendation for future research is to collect a larger dataset of individual longitudinal data. In this manner, individual mood levels can be predicted, and individual variations can be taken into account

Another limitation of this study is the method of analysis. In various earlier studies, it became clear that past activities and experiences can affect the current mood (Hollis et al., 2017). In this thesis, these types of factors were not considered. A recommendation would be to create features that describe smartphone usage of the previous two days to predict the current mood. Additionally, neural networks could be a suitable method to create personalized models that do not require large datasets (Jacques et al., 2017). However, this type of analysis is not currently within the capability of the author. It is recommended that future research apply neural networks to perform personalized mood predictions when individual longitudinal data is scarce.

5.4 Conclusion

It was investigated to what extent daily stress and anxiety levels of smartphone users could be predicted by analyzing smartphone usage data. The study consist of two prediction tasks and two types of models. The generic model is the first type of model that uses all the data in the dataset. In the second type of model, a group personalized model takes into account individuals by using only individuals with a similar degree of mood variation. In both models, prediction task 1 and 2 were performed. The first prediction task examined to what extent daily stress and anxiety levels of smartphone users can be predicted by only analyzing smartphone usage data. The second prediction task examined to what extent daily stress and anxiety levels can be predicted by analyzing smartphone usage data and the context when using a smartphone.

The results suggest that stress could not be predicted with smartphone usage data, and anxiety could be predicted moderately with the random forest algorithm and SVM. Adding external factors to the model improved the prediction performance of the random forest and SVM. Overall, the generic model predicting anxiety performed better than the generic model predicting stress. A reason may be that stress has more day to day fluctuations, whereas anxiety is a more steady-state feeling. Therefore, the models would probably be more accurate at predicting anxiety. Moreover, the group-personalized models do not improve prediction tasks 1 and 2. In conclusion, the models showed poor performance in predicting stress and anxiety. A reason might be that all the used features did not correlate with the two target variables.

6. References

Abbas, N.M. (2019, September 5). Machine learning and Mental Health. Retrieved from https://towardsdatascience.com/machine-learning-and-mental-health-7981a6001bd5

Ahn, C., & Ahn, H. (2018). Application of Random Over Sampling Examples (ROSE) for an Effective Bankruptcy Prediction Model. *The Journal of the Korea Contents Association*, *18*(8), 525-535.

Altuwairiqi, M., Arden-Close, E., Jiang, N., Powell, G., & Ali, R. (2019, May). Problematic Attachment to Social Media: The Psychological States vs Usage Styles. In *Proceedings of the 13th International Conference on Research Challenges in Information Science (RCIS), Brussel, Belgium* (pp. 29-31).

Becker, D., Bremer, V., Funk, B., Asselbergs, J., Riper, H., & Ruwaard, J. (2016). How to predict mood? Delving into features of smartphone-based data.

Beyens, I., Frison, E., & Eggermont, S. (2016). "I don't want to miss a thing": Adolescents' fear of missing out and its relationship to adolescents' social needs, Facebook use, and Facebook related stress. *Computers in Human Behavior*, *64*, 1-8.

Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*, *13*(Apr-1095.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, *70*(4), 213.

Constantinides, M., Busk, J., Matic, A., Faurholt-Jepsen, M., Kessing, L. V., & Bardram, J. E. (2018, October). Personalized versus generic mood prediction models in bipolar disorder. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (pp. 1700-1707). ACM.

Van Deursen, A. J., Bolle, C. L., Hegner, S. M., & Kommers, P. A. (2015). Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender. *Computers in human behavior*, *45*, 411-420.

Elhai, J. D., & Contractor, A. A. (2018). Examining latent classes of smartphone users: Relations with psychopathology and problematic smartphone use. *Computers in Human Behavior*, *82*, 159-166.

Elhai, J. D., Dvorak, R. D., Levine, J. C., & Hall, B. J. (2017). Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. *Journal of affective disorders*, *207*, 251-259.

Elhai, J. D., Hall, B. J., Levine, J. C., & Dvorak, R. D. (2017). Types of smartphone usage and relations with problematic smartphone behaviors: The role of content consumption vs. social smartphone use. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *11*(2).

Elhai, J. D., Yang, H., Fang, J., Bai, X., & Hall, B. J. (2019). Depression and anxiety symptoms are related to problematic smartphone use severity in Chinese young adults: fear of missing out as a mediator. *Addictive behaviors*.

Ferdous, R., Osmani, V., & Mayora, O. (2015, May). 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.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, *61*, 863-905.

Ghatasheh, N. (2014). Business analytics using random forest trees for credit risk prediction: A comparison study. *International Journal of Advanced Science and Technology*, *72*(2014), 19-30.

HCO. (2019). "Stress, the 'Health Epidemic of the 21st Century.'" *Www.Hcatodayblog.Com*, 30 Apr. 2019, www.scribbr.com/mla-citation-generator/#/sources/new/mla-webpage?title=mla.forms.webpage.

Hollis, V., Konrad, A., Springer, A., Antoun, M., Antoun, C., Martin, R., & Whittaker, S. (2017). What does all this data mean for my future mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being. *Human–Computer Interaction*, *32*(5-6), 208-267.

Jaques, N., Taylor, S., Sano, A., & Picard, R. (2017, September). Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation. In *IJCAI 2017 Workshop on Artificial Intelligence in Affective Computing* (pp. 17-33).

Inoue, T., & Abe, S. (2001, July). Fuzzy support vector machines for pattern classification. In *IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222)* (Vol. 2, pp. 1449-1454). IEEE.

Kuss, D. J., Kanjo, E., Crook-Rumsey, M., Kibowski, F., Wang, G. Y., & Sumich, A. (2018). Problematic mobile phone use and addiction across generations: The roles of psychopathological symptoms and smartphone use. *Journal of technology in behavioral science*, *3*(3), 141-149.

Ling, C. X., Sheng, V. S., & Yang, Q. (2006). Test strategies for cost-sensitive decision trees. *IEEE Transactions on Knowledge and Data Engineering*, *18*(8), 1055-1067.

Liu, J., Chen, J., & Ye, J. (2009, June). Large-scale sparse logistic regression. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 547-556). ACM.

Li, G., Cai, Z., Liu, X., Liu, J., & Su, S. (2019). A comparison of machine learning approaches for identifying high-poverty counties: robust features of DMSP/OLS night-time light imagery. *International Journal of Remote Sensing*, *40*(15), 5716-5736.

Liu, Y. Y., Yang, M., Ramsay, M., Li, X. S., & Coid, J. W. (2011). A comparison of logistic regression, classification and regression tree, and neural networks models in predicting violent re-offending. *Journal of Quantitative Criminology*, *27*(4), 547-573.

L**ópez-Santamaría, L. M., Gomez, J. C., Almanza-Ojeda, D. L., & Ibarra-Manzano, M. A. (2019)**. Age and Gender Identification in Unbalanced Social Media. In *2019 International Conference on Electronics, Communications and Computers (CONIELECOMP)* (pp. 74-80). IEEE.

Lukasik, K. M., Waris, O., Soveri, A., Lehtonen, M., & Laine, M. (2019). The relationship of anxiety, stress and depressive symptoms with working memory per-formance in a large non-depressed sample. *Frontiers in psychology*, *10*, 4.

Maxhuni, A., Hernandez-Leal, P., Sucar, L. E., Osmani, V., Morales, E. F., & Mayora, O. (2016). *Stress modelling and prediction in presence of scarce data. Journal of biomedical informatics, 63, 344-356.*

Miao, J. H., Miao, K. H., & Miao, G. J. (2015). Breast cancer biopsy predictions based on mammographic diagnosis using Support Vector Machine learning. *Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Bioinformatics*, *5*(4), 1-9.

Montag, C., Lachmann, B., Herrlich, M., & Zweig, K. (2019). Addictive features of social media/messenger platforms and freemium games against the background of psychological and economic theories. *International journal of environmental research and public health*, *16*(14), 2612.

Mitchell, L., & Hussain, Z. (2018). Predictors of problematic smartphone use: an examination of the Integrative Pathways Model and the role of age, gender, impulsiveness, excessive reassurance seeking, extraversion, and depression. *Behavioral Sciences*, *8*(8), 74.

Noë, B., Turner, L. D., Linden, D. E., Allen, S. M., Winkens, B., & Whitaker, R. M. (2019). Identifying Indicators of Smartphone Addiction Through User-App Interaction. *Computers in Human Behavior*, *99*, 56-65.

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012, July). How many trees in a random forest?. In *International workshop on machine learning and data mining in pattern recognition* (pp. 154-168). Springer, Berlin, Heidelberg.

López-Santamaría, L. M., Gomez, J. C., Almanza-Ojeda, D. L., & Ibarra-Manzano, M. A. (2019). Age and Gender Identification in Unbalanced Social Media. In *2019 International Conference on Electronics, Communications and Computers (CONIELECOMP)* (pp. 74-80). IEEE.

Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing*, *16*(1), 105-114.

Palmius, N., Saunders, K. E., Carr, O., Geddes, J. R., Goodwin, G. M., & De Vos, M. (2018). Group-Personalized Regression Models for Predicting Mental Health Scores From Objective Mobile Phone Data Streams: Observational Study. *Journal of medical Internet research*, *20*(10), e10194.

Pivetta, E., Harkin, L., Billieux, J., Kanjo, E., & Kuss, D. J. (2019). Problematic smartphone use: An empirically validated model. *Computers in Human Behavior*.

Pratap, A., Atkins, D. C., Renn, B. N., Tanana, M. J., Mooney, S. D., Anguera, J. A., & Areán, P. A. (2019). The accuracy of passive phone sensors in predicting daily mood. *Depression and anxiety*, *36*(1), 72-81.

Ramyachitra, D., & Manikandan, P. (2014). Imbalanced dataset classification and solutions: a review. *International Journal of Computing and Business Research (IJCBR)*, *5*(4).

Shin, C., & Dey, A. K. (2013, September). Automatically detecting problematic use of smartphones. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (pp. 335-344). ACM.

Sonak, A., & Patankar, R. A. (2015). A survey on methods to handle imbalance dataset. *Int. J. Comput. Sci. Mobile Comput*, *4*(11), 338-343.

Tantithamthavorn, C., Hassan, A. E., & Matsumoto, K. (2018). The impact of class rebalancing techniques on the performance and interpretation of defect prediction models. *IEEE Transactions on Software Engineering*.

Takao, M., Takahashi, S., & Kitamura, M. (2009). Addictive personality and problematic mobile phone use. *CyberPsychology & Behavior*, *12*(5), 501-507.

Twenge, J.M. (2017). Have Smartphones destroyed Generation? Retrieved November 26, 2019, from https: *https://www.theatlantic.com/magazine/archive/2017/09/has-the-smartphone-destroyed-a-generation/534198/*

Vahedi, Z., & Saiphoo, A. (2018). The association between smartphone use, stress, and anxiety: A meta-analytic review. *Stress and Health*, *34*(3), 347-358.

van Velthoven, M. H., Powell, J., & Powell, G. (2018). Problematic smartphone use: digital approaches to an emerging public health problem.

Wang, J. L., Wang, H. Z., Gaskin, J., & Wang, L. H. (2015). The role of stress and motivation in problematic smartphone use among college students. *Computers in Human Behavior*, *53*, 181-188.

Appendix

PCA

Figure 1: Contribution of most valuable variables

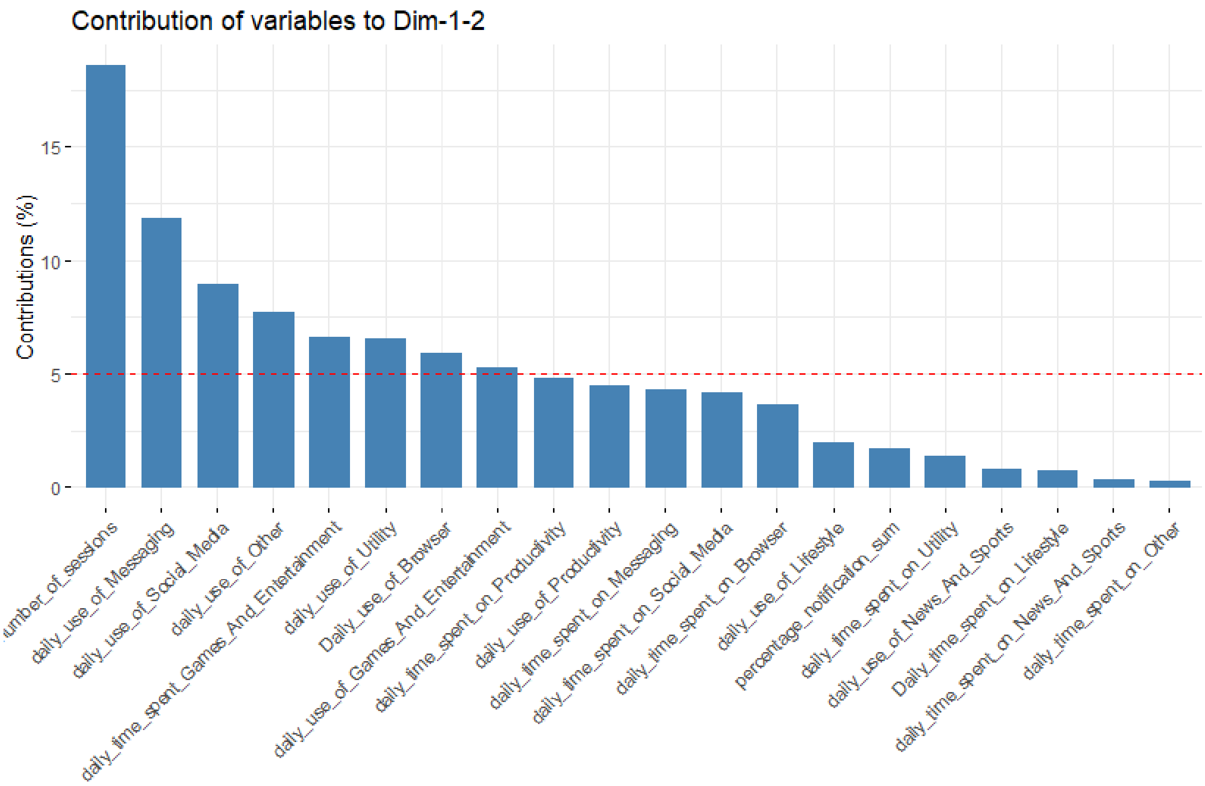

Figure 2. Variance anxiety and stress examples

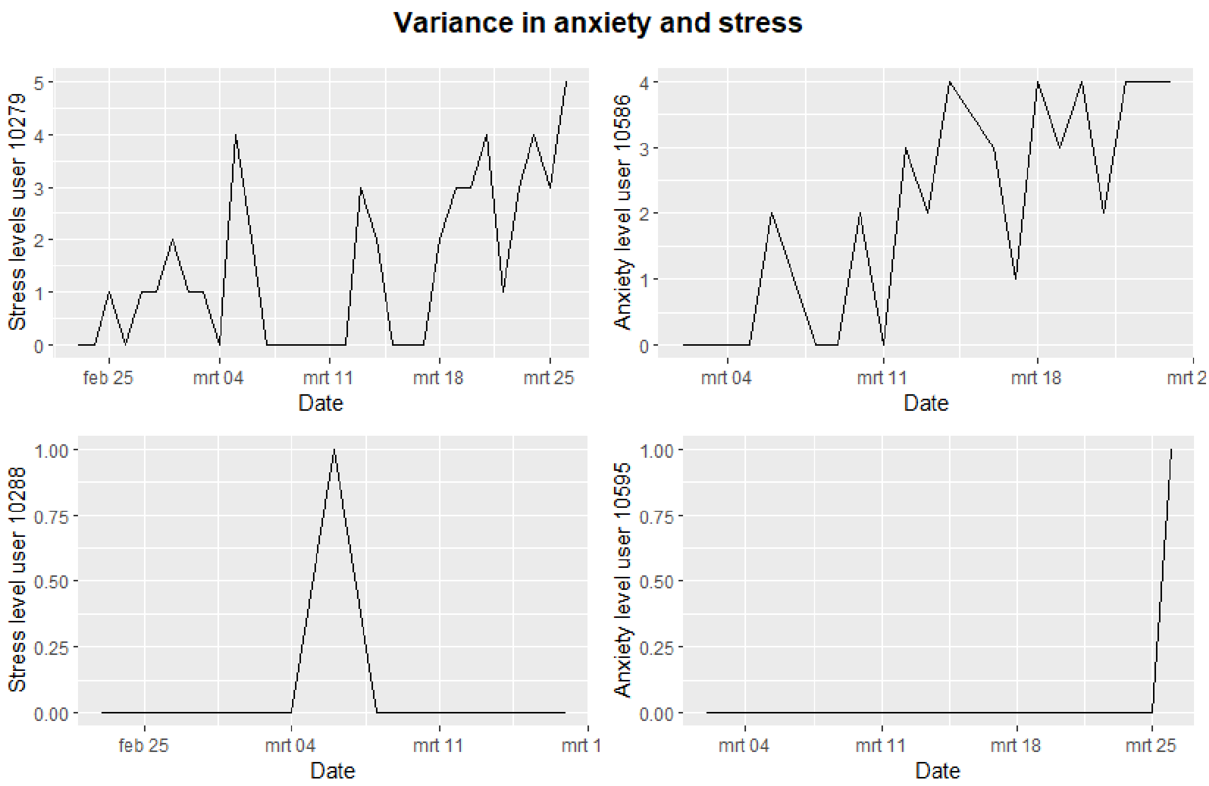

Figure 3: Correlation between smartphone features