Predicting mental tiredness among Dutch young adults by social media usage with binary classification models

C.J. Dijkhof

ANR. : U562593

SNR.: 2029730

THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY

DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICAL INTELLIGENCE

SCHOOL OF HUMANITIES AND DIGITAL SCIENCES

TILBURG UNIVERSITY

A. Hendrickson

B. Nicenboim

Tilburg University

School of Humanities and Digital Sciences

Department of Cognitive Science & Artificial Intelligence

Tilburg, The Netherlands

June 2020

Preface

This thesis was written in the context of excessive social media usage among Dutch young adults and the possible damaging results. In a society where social media and smartphones still have an increasing influence on our lives, there will be an increasing need for research containing these subjects. I hope this thesis will give insight into the effects of social media usage among young adults and show which features are most important in predicting mental tiredness.

I want to thank my supervisor for his guidance and support throughout this process during an uncommon time of our lives. I would also like to thank my fellow thesis-students, who helped and motivated me with their questions, feedback, and support. An extraordinary experience, writing and finishing this thesis during a period in our lives in which we may depend more on digital social networks than ever before due to the coronavirus.

Finally, I would like to thank my parents in particular, who supported me during this period with their love and lending a sympathetic ear.

I wish you a lot of reading pleasure.

Carien Dijkhof

Hilversum, June 26th, 2020

Abstract

In a society where young adults comprehensively use smartphones and social media, the effects of this usage are more being researched due to negative effects on well-being. This study focuses on the influence of social media usage on mental tiredness and forecasting mental tiredness among young Dutch adults by using binary classification models.

Most research done regarding the effects of social media on the mental state are from a social science or experimental origin. This research will try to predict mental tiredness based on social media usage with the use of machine learning tools. With exploratory data analysis and feature selection tools, relevant features are selected for this classification problem.

Using classification algorithms Logistic Regression, K-nearest Neighbour, Random Forest, and Support Vector Machines, this research will compare 4 main models with different subsets of data and features. The algorithms are tested on data derived from two datasets containing phone use data and self-reported mood data.

The results show an inconsistent trend against the baseline. KNN and Logistic Regression showed no clear improvement than the baseline. In general, Random Forest and SVM performed better than the baseline approaches, with Random Forest showing on average the best performance in terms of accuracy among the 4 different classification algorithms. The highest accuracy achieved was by SVM model.

The results provide new models for detecting mental tiredness among young Dutch adults, which can be used in future research. For future research, adding additional meaningful features to these models potentially improve the performance of the classification.

Keywords: Binary Classification, Mood prediction, Social media usage, Mental fatigue prediction, Mental Tiredness

Content

1. Introduction

1.1 Context

1.1.1 Social media usage among Dutch young adults.

In today's society, social media and smartphones are omnipresent. These developments generate new opportunities, such as the possibility to connect with whomever you want whenever you want, but they also result in new threats, such as smartphone- and social media addictions. In 2019, 98.5% of Dutch young adults in the age of 18 till 25 made daily use of the internet compared to 94% in the year 2012 (Central Bureau of Statistic Netherlands [CBS], 2020).

Next to the overall increase in daily use, an increase is visible in categories regarding communication platforms. Table 1 shows the most significant increases in these categories, with the highest increase in texting (e.g. WhatsApp).

Table 1

Increases in internet use by categories

Categories	2012	2019	%
Daily use	94.0%	98.5%	+4.5%
Texting	64.2%	97.7%	+33.5%
Social network sites	88.8%	92.1%	+3.3%
Professional networks	22.2%	43.6%	+21.4%

Newcom (2020)

Equivalent Dutch research among 7,000 respondents (Newcom, 2020) confirms the high use of texting app WhatsApp. The research of Newcom shows distinctive patterns in the popularity of social applications by age groups as depicted in Appendix A. It shows young adults in the age of 15 till 19 make the most use of WhatsApp, followed by YouTube, Instagram, Snapchat and Facebook. For young adults in the age of 20 to 39, WhatsApp is also the most important, but they invest more time in Facebook rather than Instagram, which has more popularity in the younger age group.

On average, Dutch young adults spend more time on social media than different age groups (Newcom, 2020). Table 2 displays the average time spent per day on social media per age group. It shows Dutch young adults between 15 and 19 spend the most time on social media, followed by young adults between 20 and 39 years. Next to age, the research shows a difference in the duration of social media use between gender and education level. The research states that women and lower educated people make more use of social media than their peers.

Table 2

Average minutes per day on social media

Age	Minutes
15-19	143
20-39	114
40-64	85
65+	82

Newcom (2020)

1.1.2 Mental tiredness.

Mental tiredness, also known as mental fatigue, can be defined as a sense of weariness (Grandjean, 1979). According to Grandjean, physical fatigue reduces muscular system performance, whereas mental fatigue causes a reduction in mental performance.
Qi et al. (2019) states continued attention during a mental-intense task, can result in high levels of mental tiredness. Mental tiredness can potentially contribute to reduced cognitive and behavioural performance such as reduced concentration, vigilance, decision making and disengagement from responsibilities (Montgomery, Montgomery, & Guisado, 1995; Qi et al., 2019.).

Research done by Ikeda and Nakamura (2014) concludes that fatigue resulting from mobile phone use is likely to be mental fatigue. As mental fatigue can influence decision making, concentration and motivation, it has a significant influence on young adults live and inter alia, their performance on educational level.

Furthermore, mental fatigue has proven to correlate with depression (Breslin, 1998; Lee, 2006). Finding out the effects of social media usage on mental tiredness can help understand mental tiredness and the prevention of it.

1.1.3 Societal and scientific relevance

In this study, the influence of social media usage on mental tiredness is analysed. Finnish research (Salmela-Aro, Upadyaya, Hakkarainen, Lonka, & Alho, 2016) concludes there is a vicious circle between burnouts, smartphone addiction and social media. By examining the prediction of mental tiredness based on social media usage on smartphones by young Dutch adults, the goal of the project is to give more insight and understanding in the effects of social media usage and possible harmful behaviour.

Prediction of mental tiredness can be very valuable since it can have considerable adverse effects on the performance of young adults and eventually can lead to burnout or depression (Breslin, 1998; Lee, 2006). The outcomes are socially relevant considering the new insight into potential unhealthy relationships of students with their smartphones and social media, which possibly can lead to a change in their potential harmful behaviour.

From a scientific point of view, this problem is worth addressing since studies relating to mental tiredness have not taken social media usage into account before (See background 2.4). Additionally, many research surrounding social media usage in combination with mental health are conducted in foreign countries (Salmela-Aro, 2016; Singh et al., 2017; Liu & Ma., 2018). In this research, the focus will be on Dutch students. Finally, this project leads to identifying features which can be used in predicting mental tiredness.

The findings of this study will redound to the benefit of society, considering social media plays a significant role in today's society. The increasing use of social media justifies the need for more insight into the effects on mental health. This research will help uncover the effects of social media on mental tiredness that have not been explored before. Thus, a new theory on the relationship between mental tiredness and social media may be arrived at.

1.2 Research questions

In this section, the research question and sub-questions are depicted. Per question, a short description is given how the question will be examined. To examine the role of social media behaviour on mental tiredness and model the prediction models, the research question is as following: “To what extent can mental tiredness be predicted among Dutch young adults, based on social media application usage?”

For answering the research question, 5 sub-questions are formulated. A short description of the answering methods is given below every question.

1) Which model predicts mental tiredness most accurate?

By creating different models based on the full dataset and relevant subsets of the dataset, the performance of each model is calculated, showing the best performing model based on accuracy.

2) Which features are most relevant in predicting mental tiredness?

By using appropriate feature selection methods, the best features for the models are selected for predicting mental tiredness among young Dutch adults.

3) To what extent does the level of mental tiredness influence the prediction? By creating 4 additional models with an adjusted binary containing only the minimum and maximum levels of mental tiredness, the influence of the extent of mental tiredness on the prediction is examined.

4) To what extent is time influencing mental tiredness?

By creating *full-day* models, the influence of time is examined. These models contain entries of 24 hours instead of 4 hours. Additionally, a feature set containing time variables is trained to examine performance with and without certain time-elements.

5) To what extent is receiving notification influencing mental tiredness?

The influence of notifications is tested with the training of an additional feature set excluding notification features.

1.3 Findings

The study shows an inconsistent trend against the baseline. KNN and Logistic Regression showed no definite improvement than the baseline. In general, Random Forest and SVM performed better than the baseline approaches, with Random Forest showing on average the best performance in terms of accuracy among the 4 different classification algorithms.

Nevertheless, the best prediction performance in terms of accuracy is obtained with the model *top25.2* model using the SVM algorithm. The models containing only the minimum and maximum levels of mental tiredness in the binary classification showed consistent better performance compared to the models with the regular binary. The *Full-day* models containing 24-hour entries showed an improvement of accuracy against the similar models (*Full)* containing 4-hour entries.

For future research, adding additional meaningful features to these models potentially improve the performance of the classification.

1.4 Outline

In chapter 2, relevant literature regarding the research problem is reviewed. The methods and the experimental setup is presented in chapter 3. In chapter 4, the results of the study are given. The limitations and recommendations for further research are presented in chapter 5. To conclude, the conclusion of the research is given in chapter 6.

2. Background

In this chapter, relevant literature regarding the research problem is outlined. In the first section of this background, multiple relevant subjects regarding social media are discussed. At the end of this chapter, relevant studies regarding the prediction of mood and mental tiredness are reviewed.

2.1 Social media usage and sleep deprivation

Social media and smartphones have a significant influence on the lives of young adults. Multiple studies show a negative influence of social media usage on sleep (Woods & Scott, 2016; Levenson, Shensa, Sidani, Colditz, & Primack, 2016, 2017). American research regarding the usage of social media in the 30 minutes before bed (Levenson et al., 2017) show a significant linear trend between the frequency of checking social media in the 30 minutes before bed and increased sleep disturbance. Åkerstedt et al. (2004) identify disturbed sleep as an essential factor in the prediction of fatigue.

2.2 Social media usage and addiction

In extreme cases, the extensive usage of social media and smartphones can result in addiction. In 2017 28.7% of young Dutch adults in the age of 18 till 25 considered themselves addicted to social media, making them the most addicted of all age-groups (CBS, 2018). Studies prove that smartphone addiction is a significant predictor of burnout and can transform into depressive symptoms (Salmela-Aro et al., 2016; Liu & Ma, 2018).

One of the main reasons young adults seem to be addicted is the fear of missing out. Franchina, Vanden Abeele, van Rooij, Lo Coco and Marez (2018) refer to the fear of missing out as feelings of anxiety arise from the realisation you might be missing out events that others are experiencing. Young adults who are more afraid to miss out, spend in general more time on social media (Buglass, Binder, Betts, & Underwood, 2017). This extra spent time leads on its turn to further feelings of missing out. Thomée, Dellve, Härenstam and Hagberg (2010) note notifications can also give a contribution to an increase of the fear of missing out.

2.3 Further effects of social media

Excessive use of social media can have great effects on mental health, such as motivation reduction, loss of interest, concentration problems, fatigue, loneliness and even depression (Aalbers, McNally, Heeren, de Wit, & Fried, 2019). Studies suggest extreme use of smartphones and social media can result in a decrease in performance on school and work, and lower academic performance (Alt, 2017; Lau, 2017).

Furthermore, multiple studies indicate a relation between social media usage and the negative influence on physical and mental well-being (Ohly & Latour, 2014; Woods & Scott, 2016; Singh, Amri, & Sabbarwal, 2017). Hunt et al. (2018) conclude limiting social media usage result in a direct positive impact on subjective well-being over time, emphasising on a likely decrease in feelings of loneliness and depression.

2.4 Mood prediction in previous studies

In the last few years, studies concerning mental health and phone use data are growing. Mood prediction based on phone use is conducted in many ways. The majority of studies concerning mental health and phone use data are based on self-reported data (Thomee, 2018). Thomee claims due to self-reported data, the outcomes may be subject to misclassification and bias. This is in line with the study conducted by Ma, Xu, Bai, Sun and Zhu (2012). They indicate that self-reported data cannot serve as a reliable indicator of objective mood without complex validation and process due to its subjectivity.

Mood prediction in combination with phone use data, commonly makes use of features counting the frequency and the duration of mobile phone data. (Falaki et al., 2010; De Montjoye et al., 2013; Becker et al., 2016; Thomee, 2018). Thomee (2018) shows most studies in the area of mood prediction include the frequency and duration spent on different applications, general screen time and time slots containing what time the phone was used.

A large number of existing studies in this area are conducted among university students. This compromises generalisability of the results. The majority of prior research conceptualised their search question as a binary classification problem. According to Thomee, most prior studies in the area of mood prediction made use of a minimum activity threshold.

Relevant studies included data about personal demographics such as age, gender, and education in variables (Falaki et al., 2010; De Montjoye et al., 2013; Becker et al., 2016; Thomee, 2018).

Furthermore, the prediction of mood with smartphone data is often executed with algorithms from machine learning and statistical modelling. Among prior studies, high diversity in algorithms is visible (Thomee, 2018). Logistic Regression (Jaques et al., 2015; Kawash, Agarwal, & Özyer, T., 2017), Random Forest (Jaques et al., 2015; van Breda et al., 2016; Meinlschmidt et al.,2020), Support Vector Machines (Jaques et al., 2015; van Breda et al., 2016., Wahle, 2016; Kawash et al., 2017) and K-Nearest Neighbour(Jaques et al., 2015; Islam et al., 2018) are widely used in the prediction of mood.

In the prediction of mental tiredness, phone data related to social media has not been used before. Prediction of mental tiredness is commonly achieved by using models with EEG-data (Trejo et al., 2005; Peng, Bouak, Wang, Chow, & Vartanian, 2018; Song, Ding, & Song al, 2019).

3. Experimental setup

This section describes the procedures that are done to obtain the results. First, a description of the used datasets is given, followed by the cleaning process. Next, the exploratory data analysis and used features are discussed. The section ends with information covering the algorithms and evaluation methods that are used during this study. In this report, data from two datasets are used containing phone use data and self-reported mood data. The study design was approved by the ethics board at Tilburg School of Humanities and Digital Sciences.

3.1 Data

3.1.1 Dataset description.

Phone_use_data.csv.

The phone use dataset (phone_use_data.csv) is collected using MobileDNA from Gent University. The dataset contains information from 124 students gathered between February 21th 2019 and March 26th 2019. The dataset is constructed with 586,792 rows of data containing information about 9 different variables. Each row contains the data of one session. The variables contain information about the id of the phone user, the application they used and the time spent on the application. Furthermore, the battery level of the phone is recorded, the session number and a binary variable are included containing information whether the user received a notification before using the used application.

Mood_sampling_data.csv.

The mood sampling dataset (mood_sampling_data.csv) contains information about 149 students gathered between June 4th 2018 and May 14th 2019. During this period, a survey was sent four times a day via the mobile application Ethica. The dataset consists of 16,016 rows of data. Each row contains the information of one survey. The data that is collected per survey is the user id, the time the survey was sent and submitted by the user and the time spent filling out the survey. Furthermore, the dataset contains information about the students' mood, activities and social contacts during a specific time frame.

3.1.2 Pre-processing and data cleaning.

Before merging the datasets into one dataset which can be used for the classification models, the data has to be pre-processed and cleaned. First, the survey entries in the mood dataset who have a non-valid output (expired, cancelled, blocked, unknown) in the duration variable are omitted. In the mental tiredness variable, some outliers are detected. The outliers are erroneous entries which exceeded the used Likert scale in the mood variables. Therefore the entries with outliers in the mental tiredness variable are eliminated.

In order to combine the data of the two datasets, the data needs to be from the same window of time. Therefore, the mood surveys which are not submitted during the time the phone use data is collected, are left out. At last, participants from the mood dataset with no registration of their phone activities are also omitted.

Research of DeLuca (2005) shows that a specific time frame can be helpful in monitoring changes in fatigue. There may be a diurnal pattern which indicates some symptoms can be more present at particular times of the day (Schwartz, 1996). Therefore, a split is made in the “response time” variable from the mood-dataset to create two new time-variables: hour and time. With the new hour variable 6 time windows of 4 hours are generated: morning, noon, afternoon, evening, night and late-night. The same is done for the phone-dataset, with the variable “end-time”. Due to the fact some surveys are submitted in quick succession, the results in each window are taken together, and the median of the variable mental tiredness is calculated for measuring. After cleaning and pre-processing, the mood dataset is reduced from 16,016 to 6,970 entries.

Social network sites and total variables.

Before the two datasets are merged, new variables are derived from the phone-use dataset. First, the number of total sessions, the total time spent and notifications per user-id during the 4-hour time window are computed. For more accurate statistics, the data of the Ethica application is removed. Subsequently, the same variables (sessions, time and notifications) are created for a subset of the phone-use data, containing only data of social network sites (SNS) sessions. In addition, the minimum, maximum and median of the time spent on social network sites is calculated. For creating the SNS subset, the data of the following social network sites were taken into account: Instagram, Facebook, YouTube, Snapchat, Twitter, LinkedIn, Pinterest and TikTok. In this research, WhatsApp is considered a direct messaging app, and not a social media platform. Therefore, the data of WhatsApp is not taken into account for the SNS variables.

Figure 1 shows the frequency in the data of the used social network sites applications. Instagram is most used, followed by Snapchat, Facebook and YouTube. From the data of these applications, additional variables are created. These variables count the number and total duration of sessions of each application during a time window.

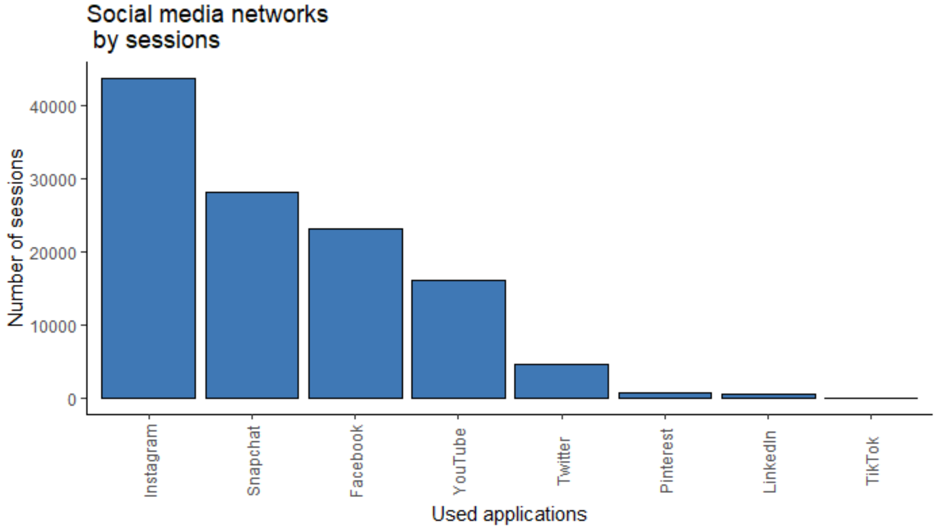

Figure 1 Frequency used applications

Time variables.

After creating the additional variables, the two datasets are merged. After merging, a binary variable is made depicting the weekend, for analysing the possible effects of the weekend on mental tiredness. In tables 3 and 4, the variables time and weekend are illustrated. After merging, the late-night window did not contain any data. Therefore, this time label is left out.

Table 3

Entries per time-window

Label	Time window	Freq.
Morning	6.00 – 9.59	224
Noon	10.00-13.59	1,426
Afternoon	14.00-17.59	1,299
Evening	18.00-21.59	1,247
Night	22.00-1.59	36

Table 4

Entries by weekend

Label	Code	Freq.
Weekday	0	2,971
Weekend	1	1,261

To increase the predictive performance of the data, the entries of users with less than 10 entries are omitted. Furthermore, sessions shorter than 1 minute or with 1 or fewer sessions are also omitted. After merging the full dataset contains 4,232 entries.

Mental tired variable.

This research focuses on the prediction of the variable mental tiredness. The variable shows positive correlations with negative mood-states and negative correlation with positive mood-states. (See appendix B). The following survey-question measures the emotion mental tiredness in the mood dataset: “I feel mentally tired (mentaal vermoeid)”. A 6 point Likert scale gives the possible replies: “Not at all”, “Very slightly”, “A little”, “Moderately”, “Quite a bit” and “Extremely”. Due to the fact self-reporting emotion and the Likert scale are both susceptible for bias (Friedman, Herskovits & Pollack,1994; Ciuk, Troy, & Jones, 2015), binary classification is generated based on the mental tiredness variable from the mood-dataset. The binary classification, depicted in figure 2, is divided by the occurrence of mental tiredness. Where class 0 ("Not at all") is compared against classes 1 ("Very slightly") to 5 ("Extremely"). Figure 3 shows the variability distribution of the emotion per respondent.

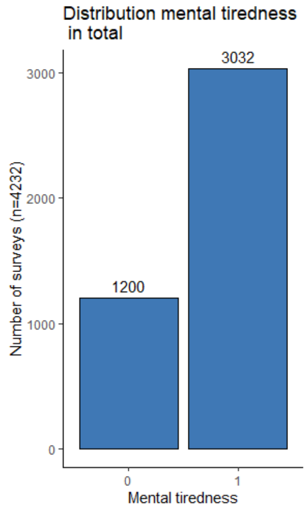

Figure 2 Mental tiredness in total

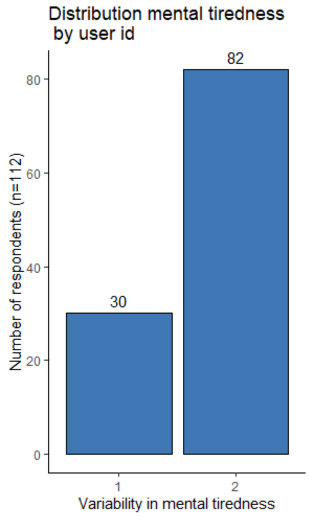

Figure 3 Mental tiredness by user_id

To answer to what extent the height of mental tiredness influences the prediction, an adjusted binary is created. This binary contains only data of the mental tiredness levels 4 and 5 (highest mental tiredness) in the positive class against the regular negative class. Table 5 shows the general distribution of the variable by the two classes in the altered *full* dataset.

Table 5

Distribution mental tiredness in the adjusted dataset

Mentally tired	0	1(4-5)
frequency	1,200	673
%	64.1%	35.9%

3.1.3 Feature selection.

To improve the generalisation and the interpretation, this study uses feature selection. With feature selection, irrelevant features are reduced, making the models easier to interpret, lowering the running time of the algorithms and creating a more general classifier (Tang, Alelyani, & Liu, 2014). For selecting the features, multiple feature selection methods are used. Prior studies used different methods for feature selection, such as experimentally removing features, stepwise regression, k-best features (Thomee,2018). Therefore, this study makes use of multiple feature selection methods. First, the correlation between the variables is calculated (See appendix B) to eliminate possible highly correlated variables. Looking at a cut point of 0.8, none of the created features need to be removed.

Learning Vector Quantization.

With Learning Vector Quantization (LVQ), the importance of the features is calculated. As depicted in figure 4, the variable insta_sessions showed to be most important, followed by insta_time. The three least important features (yt_time, sns_min_time and yt_sessions) are omitted.

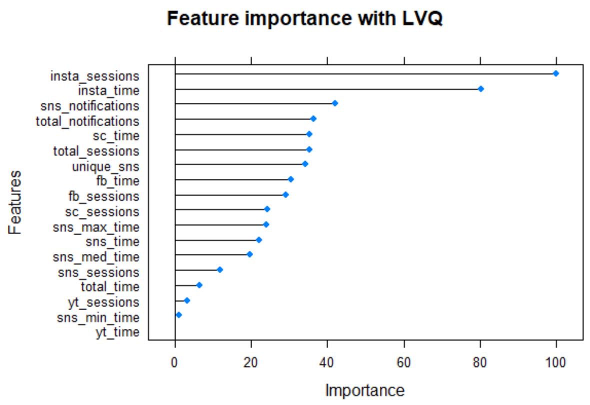

Figure 5 Feature importance LVQ

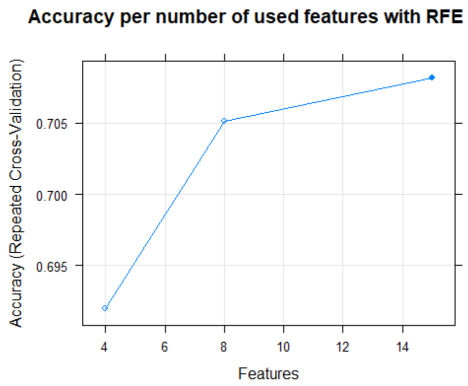

Figure 4 Accuracy features RFE

Recursive Feature Elimination

With Recursive Feature Elimination (RFE), the most relevant features in predicting the target variable are selected. The top 5 variables according to RFE, are the variables total_sessions, insta_sessions, insta_time, sc_time and sns_time. As depicted in figure 5, RFE shows the accuracy is improving with the number of features.

Akaike Information Criteria.

With Akaike Information Criteria (AIC), a stepwise feature selection is made. With stepwise feature selection, a combination is made between forward- and backward selection. The statistic fits multiple models with different sets of features, removing and adding features until the model finds its best AIC statistics (Kuhn & Johnson, 2019). After conducting AIC, the following set of features gives the best AIC statistics: sns_sessions, sns_notifications, unique_sns, total_sessions, insta_time and insta_sessions.

Final features.

After using the feature selecting methods, the final feature set (see table 6) is generated containing the most important features for predicting mental tiredness.

Table 6

Description used features.

Variable	Description	Measure
total_sessions	Total number of sessions	Count
total_time	Total time spent on the phone	Minutes
sns_sessions	Number of social network sites sessions	Count
sns_time	Time in minutes on social network sites	Minutes
Insta_sessions	Number of Instagram sessions	Count
Insta_time	Time spent on Instagram	Minutes
Sc_sessions	Number of Snapchat sessions	Count
Sc_time	Time spent on Snapchat	Minutes
Unique_sns	Unique used social network sites	Count
sns_notifications	Number of social network sites notifications	Count
total_notifications	Number of total notifications	Count

For examining the importance of time and notifications, two additional feature-sets are generated. One feature-set excluding notification-features and one feature-set including additional categorical “time” features. In table 7, the additional categorical features are presented.

Table 7

Description of additional categorical features.

Variable	Description	Measure
Weekend	*Binary if weekend*	Binary: 0-1
Time	*Time window of entry*	Multiclass: Morning, Noon, Afternoon, Evening, Night

3.1.4 Dealing with imbalanced data

The classes in this report are not equally represented. Therefore, Synthetic Minority Over-sampling Technique (SMOTE) is applied for regulating the imbalanced classes. SMOTE creates a synthetic example of the minority class with the use of k nearest neighbour (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). By selecting a random example of the minority class, the K nearest neighbours are taken, and a randomly neighbour is chosen to create a synthetic example at a randomly selected point between the two examples in feature space.

3.1.5 Cross-validation

For validating the models, most prior studies predicting mental health used k-fold cross-validation. Therefore, this study makes use of prediction models using a 10 fold cross-validation. In order to test the performance of the models, the data for the models in both R and Python are split according to the Pareto principle (Newman, 2004) in two subsets: 80% training data and 20% testing data. For tuning the parameters, the prediction models conducted in Python using the Support Vector Machines and Random Forest made use of GridSearchCV, conducting cross-validation with 5 folds.

3.2 Models & algorithms

In this section, the algorithms and models used in this study are described. The algorithms that will be used in this study are the supervised learning models K-nearest neighbour, Logistic Regression, Support Vector Machines and Random Forest. These algorithms are suitable for binary classification and frequently used similar studies (see Background 2.4).

3.2.1 K-nearest Neighbour.

K-nearest Neighbour is a basic non-parametric classification algorithm with whom the prediction boundaries are defined by K (Hastie Tibshirani, & Friedman, 2009). The algorithm calculates the distance between the training and test data. The output of the algorithm is a class membership, based on the class membership of the K nearest neighbours. The distance between the training- and test- data are calculated with distance metric Euclidean distance:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} d(x_i -, y_i)^2} \tag{1}$$

With the increase of K, the boundary becomes smoother. Subsequently, the smaller the K, the more overfitted a model gets. In this study, KNN is modelled in R. With a tuning length of 10, the algorithm picks the most suitable K for every model independently. The K which generates the highest accuracy value, is selected.

3.2.2 Logistic Regression.

Logistic regression is a classification algorithm that is commonly used to predict a binary classification. The algorithm looks for the decision boundaries among the classes. With these boundaries, the examples are split into the classes. Logistic regression has a high similarity with linear regression. The curve in logistic regression is constructed using the natural logarithm of the probabilities of the output variable (Hastie et al., 2009). The decision boundaries in logistic regression are assumed to be linear. The parameters of the model are the weights of the features. Each weighted feature is mapped into a value between 0 and 1. This value is interpreted as the probability of an example belonging to a particular class (Brid, 2018). The algorithm tunes the weights to classify the training examples correctly. Because the model assumes linear decision boundaries, the model gets less overfitted. The formal definition of logistic regression is presented below:

$$\log(\frac{P_i}{1-P_i}) = \beta_0 + \ \beta_1 X_1 + \ \beta_1 X_1 + \cdots \ \beta_n X_n + \ \varepsilon$$

Where p is the probability of the outcome, range 0 to 1. β_0 *is the intercept,* β_n *are the coefficients and* X_n *are the variables.*

In this study, the logistic regression algorithm is modelled in R. Using the GLM method in train with binomial(link = logit).

3.2.3 Support Vector Machines.

Support Vector Machines is a supervised learning algorithm commonly used for binary classification. SVM search for a hyperplane in N-dimensional space that can distinctly classify the data points. The optimal separation hyperplane is defined as the one that maximises the area between the nearest points of the class (Hastie et al., 2009). It is recommended that before using SVM, the data is standardised as the optimal separation hyperplane is influenced by the scale of the input features (Sotelo, 2017).

SVM algorithms make use of kernels, which are mathematical functions which transform the input data in the required form. In this study, radial basis function (RBF) and sigmoid are used.

Radial Basis Function:

$$K(x, x') = \exp(-\gamma \ \|x, x'\rangle\|^2) \tag{3}$$

Sigmoid:

$$K(x, x') = \tanh(k_1\langle x, x'\rangle + \ k_2) \quad (4)$$

Next to the type of kernel, SVM has two more parameters: Parameters C and Gamma. Parameter C regularise the data so that the classifier can make a better generalisation on the test data. The value of C is generally set high, which leads to slightly overfitted models. Parameter gamma defines the reach of the influence of a single training example. Low values of gamma result in farther and more influence. High values of gamma result in closer and smaller influence.

In this study, SVC from Scikit-learn was applied for the SVM. For each model, the best parameters C, gamma and kernel were chosen respectively via GridSearchCV (see appendix C).

3.2.4 Random Forest.

The Random Forest algorithm is based on a collection of multiple decision trees. The algorithm uses bagging to build an extensive collection of trees and then averaging the scores. (Hastie et al., 2009)

Multiple parameters can be tuned in a random forest. Finding the equality of the dataset can be done by Gini impurity or entropy. Gini impurity is the probability of incorrect classification if we randomly labelled according to the distribution in a branch. Entropy calculates the information gain by making a split.

Gini Impurity:

$$\sum_{i=1}^{C} f_i(1 - f_i) \quad (5)$$

Entropy:

$$\sum_{i=1}^{C} -f_i \log(f_i) \quad (6)$$

Where f_i us the frequency of label I at a node and C is the number of unique labels.

In this study, the RandomForestClassifier from Scikit-learn was used to generate the random forest models. The parameters (n_features, min_sampling_split and criterium) are tuned with GridSearchCV. In appendix C, the used parameters per model are depicted.

With minimal sampling split the model can guarantee multiple samples inform each decision. A small minimal sampling split will result in overfitting, while a large number will decrease the chances of a tree learning from the data.

3.2.5 Model implementation.

In this study, 4 different sets of the dataset are used: The *full* dataset, the *top25* containing 25 participants with more than 60 entries, the *top20* containing the top 20 participants with the most time and sessions spend on social network sites and the *full-day* dataset containing entries of 24 hours instead of 4 hours. Table 8 shows the used models in this study.

Table 8

Description used models

Model	Description	Entries	Majority Class
Full	*Entire dataset*	4,232	Tired
Full.2	*Entire dataset with adjusted binary*	1,873	Not tired
Top20s	*Dataset with top 20 most social media users*	1,403	Tired
Top20s.2	*Dataset with top 20 most social media users with adjusted binary*	750	Not tired
Top25	*Dataset with top 25 users with most entries*	1,829	Tired
Top25.2	*Dataset with top 25 users with most entries with adjusted binary*	983	Not tired
Fullday	*Dataset with 24-hour entries*	1,952	Tired
Fullday.2	*Dataset with 24-hour entries with adjusted binary*	834	Not tired

With the models containing a subset of the full dataset, the performance of the general models can be compared to more specific-user models. With these comparisons, the importance of minimal entries and certain specifics can be evaluated. With the comparison between the *full* datasets and the *full-day* datasets, the importance of the length of the window of time can be examined.

From all of these sets of data, additional sets are created which use the adjusted binary (see 3.1.2) to answer to what extent the level of mental tiredness influences the prediction. This binary contains only data of the mental tiredness levels 4 and 5 (highest mental tiredness) in the positive class against the regular negative class.

For every model, three different feature sets are used for examining the role of categorical time variables and notification variables.

3.3 Method implementation

The coding for this study is done in both R and Python. Data analysis and creating visualisations of the data are done in R (version 3.6.3). Additionally, the prediction models using the algorithms Logistic Regression and KNN are developed in R. The prediction models for the algorithms Random Forest and SVM are created using Python (3.7.6). Table 9 shows the used packages during this study with the related source and version.

Table 9

Features phone_use_dataset.csv

Used packages	Source	Version
dplyr	(Wickham, Francois, Henry, & Muller, 2020)	0.8.5
tidyr	(Wickham & Henry, 2020)	1.02
ggplot2	(Wickham, 2016)	3.3.0
caret	(Kuhn, 2020)	6.0-86
factoextra	(Kassambara & Mundt, 2020)	1.07
corrplot	(Wei & Simko, 2017)	0.84
forcats	Wickham (2020)	0.5.0
MASS	Venables, Ripley(2002)	7.3-51.5
Pandas	(McKinney, 2010)	1.0.1
Numpy	(Travis, 2006; Walt, Colbert, & Varoquaux,2011)	1.18.1
Scikit-learn	(Pedregosa, 2011)	0.22.1

3.4 Evaluation methods

In this section, the evaluation methods are presented. Evaluation of the methods has two main goals. With evaluation, the effectiveness of an algorithm can be measured. Additionally, through evaluation methods, the performance against other algorithms can be compared (Junker, Hoch, & Dengel, 1999).

3.4.1 Accuracy.

The models created for this project are optimised to gain the highest accuracy. Accuracy is a much-used evaluation metric which indicates the performance of the model. Prior studies in the area of mood prediction commonly use accuracy for evaluating the models (See Background 2.4). Therefore, this study uses accuracy for evaluating the models in such a way the performance can be evaluated and benchmarked by other research. Accuracy calculates the fraction of correctly predicted samples with the following formula:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{All\ samples} \tag{7}$$

3.4.2 Balanced accuracy.

Considering the used datasets consist of imbalanced classes, focusing merely on accuracy can be misleading. Therefore, the balanced accuracy is also calculated. With the balanced accuracy, the comparison of performance across the different datasets can be improved.

The metric for balanced accuracy uses the sensitivity of a model plus the specificity of the model divided by 2. The formal definition of balanced accuracy is depicted below.

$$Balanced\ accuracy = \frac{Sensitivity + specitifity}{2} \tag{8}$$

Next to balanced accuracy, the imbalanced classes the F1 score metric is considered. F1 scores give a weighted average of precision and recall. Due to the fact, the outcome of the F1-score is relatable to the balanced accuracy it is not taken into account.

3.4.3 Baseline.

For measuring the performance of the models, a baseline is provided. The baseline shows the accuracy when the algorithm predicts only the majority class. This baseline is also called the zero rule algorithm.

$$baseline = \frac{Majority\ class}{Minority\ class + Majority\ class} \tag{9}$$

The algorithms modelled in Python use a different method of data partition than the algorithms modelled in R. Therefore, the baseline differences between the used programmes.

4. Results

This section reports the findings of the study divided into three parts. First, the results of the exploratory data analysis are described. Thereafter, the performances of the different models and algorithms are presented.

4.1 Results exploratory data analysis

4.1.1 Distribution of mental tiredness

The difference in time and the presence of mental tiredness is presented in figure 6 and 7. As can be seen in figure 6, the results of the exploratory data analysis show that people tend to feel more mentally tired during the week than weekends. Figure 7 shows the distribution of mental tiredness by timeframe. It depicts people tend to feel more mentally tired during the evening, but less in the night. Both figures show little variance between the different timeframes.

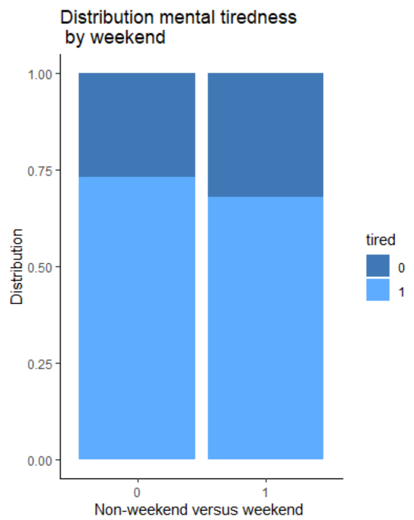

Figure 7 Distribution mental tiredness by weekend

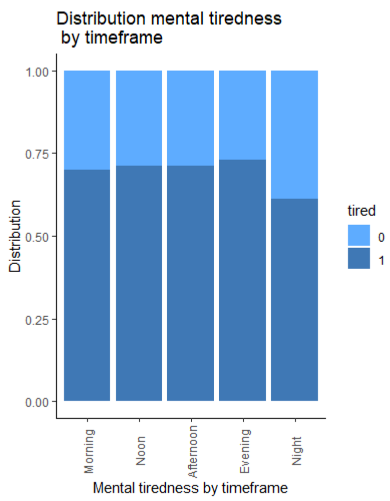

Figure 6 Distribution mental tiredness by timeframe

The results depicting the number of total notifications show people who tend to feel mentally tired, receive more notifications. As may be seen below in figure 8, the average received notifications show little variance between the perceived mental tiredness levels.

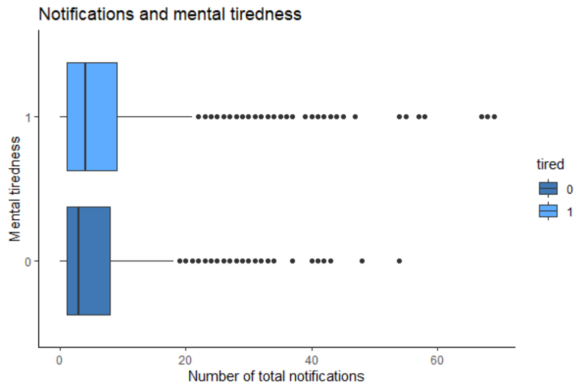

Figure 8 Notification and mental tiredness

4.1.2 Feature selection

The results show a positive correlation between the used features in this study, indicating a positive relationship. In Appendix B., the statistics of the correlation are given. For selecting the used features, multiple feature selection methods are executed. For conducting the AIC, a generalised linear model is considered. Multiple coefficient estimates showed a significant main effect for mental tiredness, as depicted in table 10 below. The results show people who tend to feel more tired, significantly make more use of social network sites (*P*<.001) and specifically Instagram (*P*<.001). Furthermore, a significant relationship is visible in the perceived mental tired levels and the received notifications from social network sites (*P*<.01), the number of used social network applications (*P*<.01), the time spend on Instagram (*P*< .01) and the total amount of sessions conducted on their phone (*P* < .01).

Table 10

Significant features.

	Estimate	Standard Error	P value
Intercept	0.604669	0.103955	P < 0.001***
sns_sessions	-0.042461	0.009022	P < 0.001***
sns_notifications	0.053839	0.020033	P < 0.01**
unique_sns	0.146466	0.044595	P < 0.01**
total_sessions	0.002873	0.001312	P < 0.01**
insta_time	-0.013031	0.004613	P < 0.01**
insta_sessions	0.077984	0.011521	P < 0.001***

The median statistics of the features, depicted in table 11, show similar results except for the increase in the number of sessions. The median statistics show an increase in sessions with a positive level of mental tiredness. The number of notifications received from social media networks shows no difference between mood levels looking at the median statistics. The same goes for the number of unique applications used during the timeframe.

In general, the results show an increase in frequency and duration with a positive level of mental tiredness.

Table 11

Median of the features per mood-level

Variable	0	1	1*
Total sessions**	47	49.5	53
Total_time	47.83	49.51	56.74
Sns_sessions***	10	11	11
Sns_time	17.59	16.39	16.67
Insta_sessions***	2	3	3
Insta_time**	1.7	4.02	3.45
Sc_sessions	0	1	0
Sc_time	0.	0.25	0
Unique_sns**	2	2	2
Sns_notifications**	0	0	0
Total_notifications	3	4	4

**Only mood levels 4 and 5*

4.1.3 Feature importance

The used feature selection methods pointed different features as the most important features. The results of the Learning Vector Quantization showed *insta_sessions* to be most important, followed by *insta_time, sns_notifications, total_notifications,* and *sc_time*. The results of the Recursive Feature Elimination qualified *total_sessions, insta_sessions, insta_time, sc*_time, and *sns_time* as the top 5 variables. The results of RFE show using more variables lead to higher accuracy. According to the results of the Akaike Information Criteria, the best AIC statistic is given with the use of the variables *sns_sessions*, *sns_notifications*, *unique_sns*, *total_sessions*, *insta_time*, and *insta_sessions*. The results show the features of the usage of Instagram are important in the prediction for mental tiredness, followed by features concerning sessions and duration in general and on social network sites. The results show inconsistent importance of features concerning notifications.

4.2 Performance of the different algorithms

Table 12 and 13 show the highest accuracies per model obtained from the different algorithms. The models scoring above baseline are depicted in green. The models scoring below baseline are depicted in red. On the right side of the tables, the mean performance in accuracy across the models is displayed. Below the models, the average performance in accuracy across the two given algorithms is depicted. The results show the lowest performance in accuracy across models is achieved by KNN (63%). The highest performance in accuracy across models is achieved by Random Forest (74%). The results of Logistic Regression (69%) are substantially better than KNN but considerably lower than SVM (73%) and Random Forest.

Table 12

Performance KNN and LR

Accuracy	*Full*	*Full.2*	*Top20s*	*Top20s.2*	*Top25*	Top25.2	*Fullday*	*Fullday.2*	*Mean*
Baseline	**0.7165**	**0.6417**	**0.7071**	**0.651**	**0.7644**	**0.7194**	**0.7026**	**0.6905**	
KNN	0.5662	0.5588	0.6143	0.6913	0.6795	0.7347	0.6103	0.5833	*0.63*
LR	0.6844	0.6471	0.6571	0.6913	0.7123	0.7398	0.6487	0.7143	*0.69*
Mean	*0.63*	*0.60*	*0.64*	*0.69*	*0.70*	*0.74*	*0.63*	*0.65*	

As depicted above in table 12, KNN and Logistic Regression showed no clear improvement than the baseline. KNN scored solely above baseline in its two best performing models. Logistic Regression scored only above baseline with the adjusted binary models.

As depicted in table 13 below, the models SVM and Random Forest both showed higher classification performance than KNN and Logistic Regression with a respectively average accuracy of 73% and 74% across all models. In general, both algorithms scored above baseline, with the exception for both on the model *Top25*.

Table 13

Performance SVM And Random Forest

Accuracy	*Full*	*Full.2*	*Top20s*	*Top20s.2*	*Top25*	Top25.2	*Fullday*	*Fullday.2*	*Mean*
Baseline	**0.7165**	**0.6417**	**0.7071**	**0.651**	**0.7644**	**0.7194**	**0.7026**	**0.6905**	
SVM	0.7214	0.6882	0.726	0.72	0.7295	0.8333	0.7263	0.6978	*0.73*
RF	0.7331	0.7147	0.7473	0.7733	0.7239	0.7195	0.7419	0.757	*0.74*
Mean	*0.73*	*0.70*	*0.74*	*0.75*	*0.73*	*0.78*	*0.73*	*0.73*	

The tables show KNN performs less than the other algorithms. Random Forest shows the overall best performance prediction based on accuracy, outperforming the other algorithms in most models. Considering the used dataset consists of imbalanced classes, focusing merely on accuracy can be misleading. Therefore, the balanced accuracy is also calculated. Table 14 shows the balanced accuracy of the best performing model per algorithm. In general, the performance based on balanced accuracy is in line with the performance based on accuracy.

Table 14

Balanced accuracy by model and algorithm

	KNN	Logistic Regression	SVM	Random Forest	*Mean*
Full	0.5135	0.5645	0.5398	0.5832	*0.55*
Full.2	0.5438	0.5833	0.6118	0.6349	*0.59*
Top20s	0.5829	0.6075	0.5234	0.5858	*0.57*
Top20s.2	0.6737	0.6558	0.7178	0.7613	*0.70*
Top25	0.6428	0.6630	0.5	0.5369	*0.59*
Top25.2	0.7324	0.6861	0.8311	0.6342	*0.72*
Fullday	0.5608	0.5462	0.5908	0.6054	*0.58*
Fullday.2	0.5816	0.6827	0.6533	0.6858	*0.65*
Mean	*0.6039*	*0.6236*	*0.621*	*0.6284*	

The performance of the models lies approximately between 55% and 75% with the maximum performance of 83.1% and a minimum performance of 50%. As can be seen above, the average balanced accuracy performance across algorithms is the highest for Random Forest. The table indicates the different performances of the algorithms based on balanced accuracy lie closer than the different performances based on pure accuracy.

4.3 Performances of the different models

The results in table 12 and 13 show the majority of the algorithms obtained their best performance in model *Top25.2.* Only Random Forest obtained the highest accuracy in model *Top20s.2*. The results depicted in table 14, show the highest balanced accuracy across models is also obtained with the *Top25.2.*

Table 15

Description used models

Model	Best performing algorithm *(Based on Accuracy)*	Accuracy	Balanced Accuracy
Full	Random Forest	0.7331	0.5832
Full.2	Random Forest	0.7147	0.6349
Top20s	Random Forest	0.7473	0.6075*
Top20s.2	Random Forest	0.7733	0.7613
Top25	Support Vector Machines	0.7295	0.6630*
Top25.2	**Support Vector Machines**	**0.8333**	**0.8311**
Fullday	Random Forest	0.7419	0.6054
Fullday.2	Random Forest	0.757	0.6858

**obtained with Logistic Regression algorithm*

As depicted with bold font in table 15, the highest accuracy obtained is with model *Top25.2* reaching an accuracy of 83% using SVM. The second best-accuracy score is obtained with model *Top20s.2* (77.3%) followed by the model *Fullday.2* (75.7%), both using the Random Forest algorithm.

Comparing the results of the adjusted binary models against the regular models, all adjusted binary models obtain higher values of balanced accuracy (See Appendix D).

As shown in table 16, both *Fullday* models obtain higher performance on balanced accuracy than the similar *Full* models. The only exception holds for the full model conducted with Logistic Regression.

Table 16

Balanced accuracy full and full-day models

	KNN	Logistic Regression	SVM	Random Forest	*Mean*
Full	0.5135	0.5645	0.5398	0.5832	*0.55*
Fullday	0.5608	0.5462	0.5908	0.6054	0.58
Full.2	0.5438	0.5833	0.6118	0.6349	*0.59*
Fullday.2	0.5816	0.6827	0.6533	0.6858	0.65

4.4 Performance of the different feature sets

The results of the performance based on the different feature-sets show varying outcomes. In table 17 and 18 below the highest achieves performance based on accuracy is highlighted with bold font. Looking at the accuracies acquired from Support Vector Machine, most models using SVM achieve slightly higher performance with *featureset* 2, which contains additional categorical features.

Table 17

Performance SVM

Accuracy	*Full*	*Full.2*	*Top20s*	*Top20s.2*	*Top25*	Top25.2	*Fullday*	*Fullday.2*	*Mean*
Baseline	**0.7107**	**0.6507**	**0.7331**	**0.5867**	**0.7295**	**0.6954**	**0.7161**	**0.6509**	
Featureset 1	0.6966	0.6266	0.7117	**0.72**	0.7158	**0.8333**	0.7163	**0.6978**	*0.72*
Featureset 2	**0.7214**	**0.6882**	**0.726**	0.68	**0.7295**	0.7885	0.6803	0.6923	*0.71*
Featureset 3	0.7048	0.68	0.7117	0.7133	0.7131	0.75	**0.7263**	0.6923	*0.71*

Models using Random Forest as depicted in table 18, achieve slightly higher performance with *featureset 3*. Models tried for Logistic Regression, on the other hand, show the highest performance with *featureset 3* (See appendix D). Models tried for KNN showed little favour for *featureset 1*.

Table 17

Performance Random Forest

Accuracy	*Full*	*Full.2*	*Top20s*	*Top20s.2*	*Top25*	Top25.2	*Fullday*	*Fullday.2*	*Mean*
Baseline	**0.7107**	**0.6507**	**0.7331**	**0.5867**	**0.7295**	**0.6954**	**0.7161**	**0.6509**	
Featureset 1	0.7118	**0.7147**	0.7401	0.7533	0.7131	0.7192	0.7213	**0.757**	*0.73*
Featureset 2	**0.7331**	0.704	**0.7473**	0.76	0.7021	0.6995	0.7213	0.7511	*0.73*
Featureset 3	0.7249	0.7041	0.744	**0.7733**	**0.7239**	**0.7195**	**0.7419**	0.7096	*0.73*

Looking at the average accuracy across models, *featureset 1* performs in general better than the other feature sets. Nevertheless, the results show no significant improvement with one of the feature-sets.

5. Discussion

In this section, the results are evaluated, and the limitations and suggestions for further research are given. Before answering the research question, the 4 sub-questions are discussed. The goal of this research is answering the question to what extent mental tiredness can be predicted among Dutch young adults, based on social media application usage.

5.1 Model performance

The results indicate that the model *Top25.2* has the highest prediction performance. The model showed in average the highest performance across the 4 algorithms. The results suggest models based on fewer people but with a certain level of entries gain a higher performance based on accuracy. This proposes a more user-generic prediction model will work better than general models based on all the users.

5.2 Feature performance

The features most relevant in the predicting of mental tiredness are the features counting the time and sessions of social network sites, Instagram and the total numbers. A similar conclusion was reached by Thomee (2018). It seems likely additional features may improve the prediction performance.

5.3 Mental tiredness and social media behaviour

The results show a small difference in social media behaviour between the mental tired entries and non-mental tired entries. The significant relation between mental tiredness and the features indicate people who feel mentally tired, tend to make more use of their phone and receive more notifications. This is in line with the literature, which indicates that more phone use has a negative influence on mental health (Ohly & Latour, 2014; Woods & Scott, 2016; Singh, Amri, & Sabbarwal, 2017). The relation between mental tiredness and the features become more apparent, looking at the adapted binary classification. Contrary to the literature, entries with levels of mental tiredness showed shorter time spend on social network sites than entries without feelings of mental tiredness. Apart from this slight disagreement, the results show a confirmation of a relationship between social media usage and mental tiredness.

Looking at the balanced accuracy, the models with the adjusted binary showed better performance than the corresponding models with the standard binary. These findings show that the level of mental tiredness influences the prediction. Using only the minimum and maximum levels of mental tiredness increases the performance of the prediction. This can be related to the fact that self-reported mental states are easily biased, and the levels can be interpreted differently per person.

5.4 Influence of time

Contrary to the expected outcome, the results show little influence of the time windows on mental tiredness, though literature shows mobile phone usage at specific moments can take significant effects on mental tiredness (Åkerstedt et al., 2004;Woods & Scott, 2016; Levenson, Shensa, Sidani, Colditz, & Primack, 2016, 2017). Due to the small amount of data for the night and late-night usage, little can be said about these time windows. The results show that people tend to feel more tired during the week compared to the weekend. In general, adding the categorical feature of weekends and time-windows did not increase the performance of models.

The performance of models predicting the mood based on 24 hours of data showed higher performance than the *full* model with time windows of 4 hours. This suggests prediction based on averaging the moods of one day, decreases bias and make the prediction easier. These results need to be interpreted with caution, as time windows can be helpful in more user-specific models discovering diurnal patterns for each person.

5.5 Influence of notifications

Receiving notifications show little influence on mental tiredness. On average, the increase in notifications shows an increase in mental tiredness. Moreover, the models in which excluded notification features predicted generally not below the models that included notification features.

5.6 Predicting mental tiredness

Predicting mental tiredness on social media usage is, to a certain extent, predictable based on social media usage. Though different measures are needed, and specific subsets have to be made. An important implication of these findings is that we cannot rule out that mental tiredness may have influenced social media usage itself. This is in line with previous results (Lin et al.,2016) which suggest mental-wellbeing and social media usage work in a vicious cycle. However, the results indicate a relation between social media usage and mental tiredness. These results are consistence with other studies which have shown a relationship between social media usage and the negative influence on physical and mental well-being (Ohly & Latour, 2014; Woods & Scott, 2016; Singh, Amri, & Sabbarwal, 2017). In conclusion, the results suggest reducing smartphone and social media usage will break the cycle and have positive effects on the mental tiredness levels.

5.7 Drawbacks and suggestions on further research

The findings of this study have to be seen in light of some limitations. Before using the data of the dataset, a significant amount of data was omitted. In future research, it is recommended to look if it is possible to retain or collect more valuable data.
Furthermore, the findings may not be representative to all Dutch young adults as the data collection was held between university students with a social science background in exchange for course credit. Future work should target more variance in educational background, and preferably take other sociodemographic statistics into account.

The mental tiredness variable is gathered via a 6-point Likert Scale survey. The answers to the survey can be subject to social desirability bias. Future studies on the current topic are therefore advised implementing physiological measures which are not subject to social desirability bias. Moreover, the validity of the mood-surveys can be enhanced by increasing the number of questions depicted for one mood and implementing a measuring of the general mood.

6. Conclusion

In this study, the main goal was developing a model for predicting mental tiredness among Young Dutch adults using their social media usage. This study presented multiple methods for predicting mental tiredness. Though most methods produced models that scored around or under the baseline, the models using random forest and support vector machines showed promising predictions. From the research that has been carried out, it is possible to conclude that a relationship occurs between mental tiredness and social media usage. The results suggest lowering social media usage, will improve the levels of mental tiredness. Our methods could be applied for further research, although future work will take into account predicting mental tiredness solely on social media usage is not enough. This study shows that different features also need to be taken into account. For predicting, a model containing a higher minimum threshold of entries per participant is beneficial for the prediction. With more data per person, the patterns become more visible.

This study shows that the prediction of mental tiredness based on the currently used data is challenging. In future research, adding additional meaningful features, such as sociodemographic features and physiological measures, will potentially improve the performance of the classification. This study confirms a significant relationship between mental tiredness and social media usage and gives an insight on the effects of smartphones and social media usage. Further work needs to be performed to establish the exact causal relationship between social media and mental tiredness.

References

Acikmese, Y., & Alptekin, S. E. (2019). Prediction of stress levels with LSTM and passive mobile sensors. *Procedia Computer Science*, *159*, 658–667. https://doi.org/10.1016/j.procs.2019.09.221

Åkerstedt, T., Knutsson, A., Westerholm, P., Theorell, T., Alfredsson, L., & Kecklund, G. (2004). Mental fatigue, work and sleep. *Journal of Psychosomatic Research*, 57(5), 427–433. https://doi.org/10.1016/j.jpsychores.2003.12.001

Alt, D. (2017). Students' social media engagement and fear of missing out (FoMO) in a diverse classroom. *Journal of Computing in Higher Education*, 29(2), 388–410. https://doi.org/10.1007/s12528-017-9149-x

Becker, D., & Bremer, V., Funk, B., & Asselbergs, J., Riper, H., & Ruwaard, J. (2016). How to Predict Mood? Delving into Features of Smartphone- Based Data. *Americas Conference on Information Systems, AMCIS 2016 (2016)*

Belardi, A., Stalujanis, E., Oh, M., Jung, E. K., … Lee, J.-H. (2020). Personalized prediction of smartphone-based psychotherapeutic micro-intervention success using machine learning. *Journal of Affective Disorders*, *264*, 430–437. https://doi.org/10.1016/j.jad.2019.11.071)

Breslin, E., van der Schans, C., Breukink, S., Meek, P., Mercer, K., Volz, W., & Louie, S. (1998). Perception of Fatigue and Quality of Life in Patients With COPD. *Chest*, 114(4), 958–964. https://doi.org/10.1378/chest.114.4.958

Brid, R. S. (2018, October 19). Logistic Regression. Retrieved from https://medium.com/greyatom/logistic-regression-89e496433063

Buglass, S. L., Binder, J. F., Betts, L. R., & Underwood, J. D. M. (2017). Motivators of online vulnerability: The impact of social network site use and FOMO. *Computers in Human Behavior*, 66, 248–255. https://doi.org/10.1016/j.chb.2016.09.055

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

Ciuk, D., Troy, A. K., & Jones, M. C. (2015). Measuring Emotion: Self-Reports vs. Physiological Indicators. *SSRN Electronic Journal* https://doi.org/10.2139/ssrn.2595359

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports,* 3(1) https://doi.org/10.1038/srep01376

Falaki, H., Mahajan, R., Kandula, S., Lymberopoulos, D., Govindan, R., & Estrin, D. (2010). Diversity in smartphone usage. *Proceedings of the 8th international conference on Mobile systems, applications, and services - MobiSys '10*. https://doi.org/10.1145/1814433.1814453

Forthofer, R. N., Lee, E. S., & Hernandez, M. (2007). Logistic and Proportional Hazards Regression. *Biostatistics*, 387–419. https://doi.org/10.1016/b978-0-12-369492-8.50019-4

Franchina, V., Vanden Abeele, M., van Rooij, A., Lo Coco, G., & De Marez, L. (2018). Fear of Missing Out as a Predictor of Problematic Social Media Use and Phubbing Behavior among Flemish Adolescents. *International Journal of Environmental Research and Public Health,* 15(10), 2319. https://doi.org/10.3390/ijerph15102319

Friedman, H.H., Herskovitz, P.J., Pollack, S. (1994). The biasing effects of scale-checking styles on response to a likert scale Proceedings of the American Statistical Association *Annual Conference: Survey Research Methods* pp. 477-481

Grandjean, E. (1979). Fatigue in industry. *Occupational and Environmental Medicine*, 36(3), 175–186. https://doi.org/10.1136/oem.36.3.175

Hastie, T., & Tibshirani, R. (2009). *Elements of Statistical Learning* (2nd ed. 2009, Corr. 9th printing 2017 ed.). Heidelberg, Germany: Springer Medizin Verlag.

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No More FOMO: Limiting Social Media Decreases Loneliness and Depression. *Journal of Social and Clinical Psychology*, *37*(10), 751–768. https://doi.org/10.1521/jscp.2018.37.10.751

Ikeda, K., Nakamura, K. Association between mobile phone use and depressed mood in Japanese adolescents: a cross-sectional study. *Environ Health Prev Med* 19, 187–193 (2014). https://doi.org/10.1007/s12199-013-0373-3

Islam, M. R., Kamal, A. R. M., Sultana, N., Islam, R., Moni, M. A., & Ulhaq, A. (2018). Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique. 2018 *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering* (IC4ME2). https://doi.org/10.1109/ic4me2.2018.8465641

Jaques, N., Taylor, S., Azaria, A., Ghandeharioun, A., Sano, A., & Picard, R. (2015). Predicting students' happiness from physiology, phone, mobility, and behavioral data. *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 222-228. https://doi.org/10.1109/acii.2015.7344575

Junker, M., Hoch, R., Dengel, A. (1999). "On the evaluation of document analysis components by recall, precision, and accuracy," Proceedings of the Fifth International Conference on Document Analysis and Recognition. *ICDAR '99* (Cat. No.PR00318), Bangalore, India, 1999, pp. 713-716, doi: 10.1109/ICDAR.1999.791887.

Kang, S., & Kurtzberg, T. R. (2019). Reach for your cell phone at your own risk: The cognitive costs of media choice for breaks. *Journal of Behavioral Addictions*, 8(3), 395–403. https://doi.org/10.1556/2006.8.2019.21

Kawash, J., Agarwal, N., & Özyer, T. (2017). Prediction and Inference from Social Networks and Social Media. New York, United States: Springer Publishing.

Kuhn, M., & Johnson, K. (2019). Feature Engineering and Selection. Abingdon, United Kingdom: Taylor & Francis.

Lau, W. W. F. (2017). Effects of social media usage and social media multitasking on the academic performance of university students. *Computers in Human Behavior*, *68*, 286–291. https://doi.org/10.1016/j.chb.2016.11.043

Lee, W. H., & Kim, C. J. (2006). The Relationship between Depression, Perceived Stress, Fatigue and Anger in Clinical Nurses. *Journal of Korean Academy of Nursing*, *36*(6), 925. https://doi.org/10.4040/jkan.2006.36.6.925

Lin, L. yi, Sidani, J. E., Shensa, A., Radovic, A., Miller, E., Colditz, J. B., … Primack, B. A. (2016). Association Between Social Media Use And Depression Among U.S. Young Adults. *Depression and Anxiety*, *33*(4), 323–331. https://doi.org/10.1002/da.22466

Liu, C., & Ma, J. (2018). Social media addiction and burnout: The mediating roles of envy and social media use anxiety. *Current Psychology*. https://doi.org/10.1007/s12144-018-9998-0

Ma, Y., Xu, B., Bai, Y., Sun, G., & Zhu, R. (2012). Daily Mood Assessment Based on Mobile Phone Sensing. *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks*, 1. https://doi.org/10.1109/bsn.2012.3

Max Kuhn (2020). caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*. https://doi.org/10.25080/majora-92bf1922-00a

Meinlschmidt, G., Tegethoff, M., Belardi, A., Stalujanis, E., Oh, M., Jung, E. K., Kim, H. C., Yoo, S. S., & Lee, J. H. (2020). Personalized prediction of smartphone-based psychotherapeutic micro-intervention success using machine learning. Journal of *Affective Disorders*, 264, 430-437. https://doi.org/10.1016/j.jad.2019.11.071

Montgomery, L. D., Montgomery, R. W., & Guisado, R. (1995). Rheoencephalographic and electroencephalographic measures of cognitive workload: analytical procedures. *Biological Psychology*, 40(1–2), 143–159. https://doi.org/10.1016/0301-0511(95)05117-1

Newcom. (2020, January). *Nationaal Social Media Onderzoek 2020*. Retrieved from https://www.newcom.nl/downloads/2020-NSMO-Rapportage-2020-Publicatieversie-25012020.pdf?utm_source=ActiveCampaign&utm_medium=email&utm_content=Nationale+Social+Media+onderzoek+2020

Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. Contemporary Physics, 46(5), 323–351. https://doi.org/10.1080/00107510500052444

Pedregosa, F.,Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., …, Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825—2830.

Peng, H. T., Bouak, F., Wang, W., Chow, R., & Vartanian, O. (2018). An improved model to predict performance under mental fatigue. *Ergonomics*, 61(7), 988–1003. https://doi.org/10.1080/00140139.2017.1417641

Przybylski, A. K., & Weinstein, N. (2017). A Large-Scale Test of the Goldilocks Hypothesis. *Psychological Science*, *28*(2), 204–215. https://doi.org/10.1177/0956797616678438

Qi, P., Ru, H., Gao, L., Zhang, X., Zhou, T., Tian, Y., … Sun, Y. (2019). Neural Mechanisms of Mental Fatigue Revisited: New Insights from the Brain Connectome. *Engineering*, 5(2), 276–286. https://doi.org/10.1016/j.eng.2018.11.025

R Core Team (2020). R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Singh, M. M., Amiri, M., & Sabbarwal, S. (2017). Social Media Usage: Positive and Negative

Effects on the Life Style of Indian Youth. Unpublished. https://doi.org/10.13140/RG.2.2.24483.20003

Song, A., Niu, C., Ding, X., Xu, X., & Song, Z. (2019). Mental Fatigue Prediction Model Based on Multimodal Fusion. *IEEE Access*, *7*, 177056–177062. https://doi.org/10.1109/access.2019.2941043

Sotelo, D. (2017, July 26). *Effect of Feature Standardization on Linear Support Vector Machines*. Retrieved from https://towardsdatascience.com/effect-of-feature-standardization-on-linear-support-vector-machines-13213765b812

Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. In *Data Classification: Algorithms and Applications* (pp. 37-64). CRC Press. https://doi.org/10.1201/b17320

Thomée, S., Dellve, L., Härenstam, A., & Hagberg, M. (2010). Perceived connections between information and communication technology use and mental symptoms among young adults - a qualitative study. *BMC Public Health*, 10(1). https://doi.org/10.1186/1471-2458-10-66

Thomée, S. (2018). Mobile Phone Use and Mental Health. A Review of the Research That Takes a Psychological Perspective on Exposure. International Journal of Environmental Research and Public Health, 15(12), 2692. https://doi.org/10.3390/ijerph15122692

Travis E, Oliphant. A guide to NumPy, USA: Trelgol Publishing, (2006). van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, 13(2), 22–30. https://doi.org/10.1109/mcse.2011.37

Trejo, L. J., Kochavi, R., Kubitz, K., Montgomery, L. D., Rosipal, R., & Matthews, B. (2005).

Measures and models for predicting cognitive fatigue. *Biomonitoring for Physiological and Cognitive Performance during Military Operations*. https://doi.org/10.1117/12.604286

Van Breda, W., Pastor, J., Hoogendoorn, M., Ruwaard, J., Asselbergs, J., & Riper, H. (2016). Exploring and Comparing Machine Learning Approaches for Predicting Mood Over Time. *Innovation in Medicine and Healthcare* 2016, 37–47. https://doi.org/10.1007/978-3-319-39687-3_4

Van Messem, A. (2020). Support vector machines: A robust prediction method with applications in bioinformatics. *Handbook of Statistics*, 391–466.

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M., & Weidt, S. (2016). Mobile Sensing and Support for People With Depression: A Pilot Trial in the Wild. *JMIR mHealth and uHealth*, *4*(3), e111. https://doi.org/10.2196/mhealth.5960

Wei, T., Simko, V. (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

Wickham, H. (2020). forcats: Tools for Working with Categorical Variables (Factors). R package version 0.5.0. **https://CRAN.R-project.org/package=forcats**

Wickham, H., Henry, L. (2020). tidyr: Tidy Messy Data. R package version 1.0.2. https://CRAN.R-project.org/package=tidyr

Wickham, H., François, R., Henry, L., Müller, K. (2020). dplyr: A Grammar of Data Manipulation. R package version 0.8.5. https://CRAN.R-project.org/package=dplyr

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics* *and Intelligent Laboratory Systems*, *2*(1–3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

Appendix A: Background

Appendix A1: Social media usage by relevant age-group

Table A1

Social media usage by relevant age-group

Age	15-19		20-39	
Media %	Total use	Daily use	Total use	Daily use
Facebook	51%	26%	82%	57%
YouTube	90%	53%	79%	29%
Instagram	82%	60%	57%	41%
Snapchat	72%	52%	35%	21%
Twitter	21%	9%	26%	10%
LinkedIn	9%	1%	47%	7%
Pinterest	21%	4%	32%	6%
TikTok	17%	6%	9%	4%

Newcom (2020)

Appendix B: Exploratory Data Analysis

Appendix B1: Correlation matrix mood

Table B1

Correlation matrix mood

	anxious	bored	gloomy	calm	stressed	content	cheerful	tired	energetic	upset	envious	inferior
anxious	1											
bored	0.314	1										
gloomy	0.638	0.355	1									
calm	-0.419	-0.113	-0.391	1								
stressed	0.595	0.243	0.550	-0.410	1							
content	-0.333	-0.186	-0.385	0.554	-0.323	1						
cheerful	-0.330	-0.203	-0.397	0.513	-0.321	0.679	1					
tired	**0.457**	0.331	**0.555**	-0.327	**0.554**	-0.321	-0.383	1				
energetic	-0.233	-0.198	-0.330	0.431	-0.247	0.544	0.695	**-0.416**	1			
upset	0.585	0.288	0.621	-0.351	0.522	-0.302	-0.303	**0.456**	-0.215	1		
envious	0.421	0.337	0.414	-0.181	0.342	-0.154	-0.129	0.294	-0.078	0.443	1	
inferior	0.457	0.275	0.548	-0.285	0.420	-0.264	-0.273	**0.414**	-0.223	0.491	0.549	1

Appendix B2: Correlation matrix important features

Table B2

Correlation matrix important features

	total_ sessions	total_ time	sns_ sessions	sns_ time	insta_ sessions	insta_time	sc_ sessions	sc_ time	unique_ sns	sns_ notifications	total_ notifications
total_sessions	1										
total_time	0.4685	1									
sns_sessions	0.6847	0.4434	1								
sns_time	0.232	0.733	0.4966	1							
insta_sessions	0.4975	0.2521	0.7044	0.2525	1						
insta_time	0.2305	0.3357	0.433	0.4047	0.6821	1					
sc_sessions	0.4295	0.1077	0.6071	0.1201	0.2118	0.0963	1				
sc_time	0.2916	0.1404	0.4352	0.1694	0.1487	0.1007	0.7427	1			
unique_sns	0.2692	0.2355	0.4839	0.3007	0.3426	0.2557	0.2766	0.2488	1		
sns_notifications	0.1999	0.0919	0.3774	0.086	0.2077	0.1358	0.5248	0.4014	0.2709	1	
total_notifications	0.3915	0.1546	0.3731	0.061	0.2569	0.1379	0.3396	0.2335	0.2619	0.6897	1

Appendix C Parameter optimising

Appendix C1: Results Random Forest parameter optimising

Table C1

Results Random Forest parameter optimising

Model	Featureset 1	Featureset 2	Featureset 3
1	Gini min_samples_split = 5 n_estimators = 100	Gini min_samples_split = 2 n_estimators = 200	Gini min_samples_split = 2 n_estimators = 200
1a	Entropy min_samples_split = 5 n_estimators = 200	Gini min_samples_split = 5 n_estimators = 200	Entropy min_samples_split = 5 n_estimators = 50
2	Gini min_samples_split =10 n_estimators = 100	Gini min_samples_split =5 n_estimators = 200	Entropy min_samples_split = 10 n_estimators = 100
2a	Entropy min_samples_split = 10 n_estimators = 200	Gini min_samples_split = 5 n_estimators = 200	Entropy min_samples_split = 10 n_estimators = 200
3	Entropy min_samples_split = 20 n_estimators = 200	Entropy min_samples_split = 10 n_estimators = 50	Gini min_samples_split = 20 n_estimators = 50
3a	Entropy min_samples_split = 2 n_estimators = 100	Entropy min_samples_split = 2 n_estimators = 200	Entropy min_samples_split = 2 n_estimators = 200
4	Gini min_samples_split = 5 n_estimators = 50	Gini min_samples_split = 2 n_estimators = 200	Entropy min_samples_split = 2 n_estimators = 200
4a	Entropy min_samples_split = 2 n_estimators = 200	Entropy min_samples_split = 5 n_estimators = 200	Gini min_samples_split = 2 n_estimators = 100

Appendix C2: Results SVM parameter optimising

Table C2

Results SVM parameter optimising

Model	Featureset 1	Featureset 2	Featureset 3
1	Radial Basis Function	Radial Basis Function	Radial Basis Function
	C = 10	C = 1	C = 10
	Gamma = 0.1	Gamma = 0.1	Gamma = 0.1
1a	Radial Basis Function	Radial Basis Function	Radial Basis Function
	C = 100	C = 100	C = 10
	Gamma = 0.1	Gamma = 0.01	Gamma = 0.1
2	Radial Basis Function	Radial Basis Function	Radial Basis Function
	C = 10	C = 10	C = 1
	Gamma = 1	Gamma = 0.01	Gamma = 1
2a	Radial Basis Function	Radial Basis Function	Radial Basis Function
	C = 10	C = 10	C = 10
	Gamma = 0.1	Gamma = 0.01	Gamma = 0.1
3	Radial Basis Function	Radial Basis Function	Radial Basis Function
	C = 1	C = 0.1	C = 1
	Gamma = 0.1	Gamma = 1	Gamma = 1
3a	Radial Basis Function	Radial Basis Function	Radial Basis Function
	C = 10	C = 1	C = 10
	Gamma = 0.1	Gamma = 0.1	Gamma = 0.1
4	Radial Basis Function	Radial Basis Function	Radial Basis Function
	C = 1	C = 10	C = 10
	Gamma = 1	Gamma = 0.1	Gamma = 0.1
4a	Radial Basis Function	Radial Basis Function	Radial Basis Function
	C = 100	C = 1	C = 10
	Gamma = 0.1	Gamma = 1	Gamma = 1

Appendix D Results Algorithms per Model

Appendix D1: Results algorithms model *Full*

Model 1: Full dataset, binary mood

1) Basic feature set

2) With categorical feature(s): timeframe + weekend

3) Without notification features

Table D1

Distribution output variable

Variable	Full	Train
Not tired	1,200	960
Tired	3,032	2,426
Total	4,232	3,386

Table D2

K Nearest Neighbour model 1

	Accuracy	Balanced accuracy
1	0.5366	0.4841
2	0.5638	0.5005
3	0.5662	0.5135
Baseline	0.7165	

Table D3

Logistic Regression model 1

	Accuracy	Balanced accuracy
1	0.6631	0.5421
2	0.6584	0.5414
3	0.6844	0.5645
Baseline	0.7165	

Table D4

Support Vector Machines model 1

	Accuracy	Balanced accuracy
1	0.6966	0.5398
2	0.7214	0.5281
3	0.7048	0.5393
Baseline	0.7107	

Table D5

Random Forest model 1

	Accuracy	Balanced accuracy
1	0.7118	0.5622
2	0.7331	0.5832
3	0.7249	0.5775
Baseline	0.7107	

Appendix D2: Results algorithms model *Full.2*

Model 1a: Full dataset with adjusted binary

1) Basic feature set

2) With categorical feature(s): timeframe + weekend

3) Without notification features

Table D6

Distribution output variable

Variable	Full	Train
Not tired	1,200	960
Tired	673	539
Total	1,873	1,499

Table D7

K Nearest Neighbour model 1a

	Accuracy	Balanced Accuracy
1	0.5561	0.5438
2	0.5508	0.5313
3	0.5588	0.5392
Baseline	0.6417	

Table D8

Logistic Regression model 1a

	Accuracy	Balanced Accuracy
1	0.639	0.5688
2	0.6471	0.5833
3	0.623	0.5480
Baseline	0.6417	

Table D9

Support Vector Machines model 1a

	Accuracy	Balanced accuracy
1	0.6266	0.5735
2	0.6882	0.6118
3	0.68	0.6097
Baseline	0.6507	

Table D10

Random Forest model 1a

	Accuracy	Balanced accuracy
1	0.7147	0.6349
2	0.704	0.6138
3	0.7041	0.6261
Baseline	0.6507	

Appendix D3: Results algorithms *Top20s*

Model 2: top 20 social media

1) Basic feature set

2) With categorical feature(s): timeframe + weekend

3) Without notification features

Table D11

Distribution output variable

Variable	Full	Train
Not tired	413	331
Tired	990	792
Total	1,403	1,123

Table D12

K Nearest Neighbour model 2

	Accuracy	Balanced Accuracy
1	0.6143	0.5665
2	0.6071	0.5829
3	0.5893	0.5417
Baseline	0.7071	

Table D13

Logistic Regression model 2

	Accuracy	Balanced Accuracy
1	0.625	0.5848
2	0.6286	0.5909
3	0.6571	0.6075
Baseline	0.7071	

Table D14

Support Vector Machine model 2

	Accuracy	Balanced Accuracy
1	0.7117	0.5234
2	0.726	0.5133
3	0.7117	0.4856
Baseline	0.7331	

Table D15

Random Forest model 2

	Accuracy	Balanced Accuracy
1	0.7401	0.5763
2	0.7473	0.5858
3	0.744	0.5789
Baseline	0.7331	

Appendix D4: Results algorithms *Top20s.2*

Model 2a: top 20 social media with adjusted binary

1) Basic feature set

2) With categorical feature(s): timeframe + weekend

3) Without notification features

Table D16

Distribution output variable

Variable	Full	Train
Not tired	486	389
Tired	264	212
Total	750	601

Table D17

K Nearest Neighbour model 2a

	Accuracy	Balanced Accuracy
1	0.6913	0.6737
2	0.651	0.6204
3	0.6242	0.5998
Baseline	0.651	

Table D18

Logistic Regression model 2a

	Accuracy	Balanced Accuracy
1	0.6913	0.6558
2	0.6779	0.6455
3	0.6644	0.6263
Baseline	0.651	

Table D19

Support Vector Machine model 2a

	Accuracy	Balanced Accuracy
1	0.72	0.7178
2	0.68	0.6751
3	0.7133	0.7106
Baseline	0.5867	

Table D20

Random Forest model 2a

	Accuracy	Balanced Accuracy
1	0.7533	0.7419
2	0.76	0.7462
3	0.7733	0.7613
Baseline	0.5867	

Appendix D5: Results algorithms model *Top25*

Model 3: top 25 entries

1) Basic feature set

2) With categorical feature(s): timeframe + weekend

3) Without notification features

Table D21

Distribution output variable

Variable	Full	Train
Not tired	431	345
Tired	1,398	1,119
Total	1,829	1,464

Table D22

K Nearest Neighbour model 3

	Accuracy	Balanced Accuracy
1	0.6575	0.6151
2	0.6795	0.6375
3	0.663	0.6428
Baseline	0.7644	

Table D23

Logistic Regression model 3

	Accuracy	Balanced Accuracy
1	0.7041	0.6536
2	0.7096	0.6451
3	0.7123	0.6630
Baseline	0.7644	

Table D24

Support Vector Machine model 3

	Accuracy	Balanced Accuracy
1	0.7158	0.5
2	0.7295	0.5
3	0.7131	0.4919
Baseline	0.7295	

Table D25

Random Forest model 3

	Accuracy	Balanced Accuracy
1	0.7131	0.5139
2	0.7021	0.5095
3	0.7239	0.5369
Baseline	0.7295	

Appendix D6: Results algorithms model *Top25.2*

Model 3a: Model top 25 entries with adjusted binary

1) Basic feature set

2) With categorical feature: weekend

3) Without notification features

Table D26

Distribution output variable

Variable	Full	Train
Not tired	708	567
Tired	275	220
Total	983	787

Table D27

K Nearest Neighbour model 3a

	Accuracy	Balanced Accuracy
1	0.7347	0.7324
2	0.7041	0.6890
3	0.7296	0.6790
Baseline	0.7194	

Table D28

Logistic Regression model 3a

	Accuracy	Balanced Accuracy
1	0.6888	0.6562
2	0.6837	0.6471
3	0.7398	0.6861
Baseline	0.7194	

Table D29

Support Vector Machine model 3a

	Accuracy	Balanced Accuracy
1	0.8333	0.8311
2	0.7885	0.7869
3	0.75	0.7464
Baseline	0.6954	

Table D30

Random Forest model 3a

	Accuracy	Balanced Accuracy
1	0.7192	0.6342
2	0.6995	0.5361
3	0.7195	0.6297
Baseline	0.6954	

Appendix D7: Results algorithms model *Fullday*

Model 4: Full days

1) Basic feature set

2) With categorical feature(s): weekend

3) Without notification features

Table D31

Distribution output variable

Variable	Full	Train
Not tired	582	466
Tired	1,370	1,096
Total	1,952	1,562

Table D32

K Nearest Neighbour model 4

	Accuracy	Balanced Accuracy
1	0.6103	0.5608
2	0.6077	0.5493
3	0.6103	0.5412
Baseline	0.7026	

Table D33

Logistic Regression model 4

	Accuracy	Balanced accuracy
1	0.6487	0.5462
2	0.6359	0.5420
3	0.6487	0.5437
Baseline	0.7026	

Table D34

Support Vector Machine model 4

	Accuracy	Balanced accuracy
1	0.7163	0.5248
2	0.6803	0.5346
3	0.7263	0.5908
Baseline	0.7161	

Table D35

Random Forest model 4

	Accuracy	Balanced accuracy
1	0.7213	0.5773
2	0.7213	0.5746
3	0.7419	0.6054
Baseline	0.7161	

Appendix D8: Results algorithms model *Fullday.2*

Model 4a: Full day model with adjusted binary

1) Basic feature set

2) With categorical feature: weekend

3) Without notification features

Table D36

Distribution output variable

Variable	Full	Train
Not tired	582	466
Tired	262	210
Total	834	676

Table D37

K Nearest Neighbour model 4a

	Accuracy	Balanced Accuracy
1	0.5833	0.5710
2	0.5833	0.5816
3	0.5714	0.5517
Baseline	0.6905	

Table D38

Logistic Regression model 4a

	Accuracy	Balanced Accuracy
1	0.7143	0.6817
2	0.7083	0.6827
3	0.6964	0.6582
Baseline	0.6905	

Table D39

Support Vector Machine model 4a

	SVM	Balanced accuracy
1	0.6978	0.6533
2	0.6923	0.5659
3	0.6923	0.6223
Baseline	0.6509	

Table D40

Random Forest model 4a

	Random Forest	Balanced accuracy
1	0.757	0.6858
2	0.7511	0.6812
3	0.7096	0.6305
Baseline	0.6509	

Appendix D9: Performance adjusted binary versus regular binary

Table 41

Full models

Balanced Accuracy	*KNN*	*LR*	*SVM*	*RF*
Full	0.5135	0.5645	0.5398	0.5832
Full.2	**0.5438**	**0.5833**	**0.6118**	**0.6349**

Table 42

Top 20 social media users

Balanced Accuracy	*KNN*	*LR*	*SVM*	*RF*
Top20s	0.5829	0.6075	0.5234	0.5858
Top20s.2	**0.6737**	**0.6558**	**0.7178**	**0.7613**

Table 43

Top 25 most entries

Balanced Accuracy	*KNN*	*LR*	*SVM*	*RF*
Top25	0.6428	0.6630	0.5	0.5369
Top25.2	**0.7324**	**0.6861**	**0.8311**	**0.6342**

Table 44

Full day models (24-hour)

Balanced Accuracy	*KNN*	*LR*	*SVM*	*RF*
Fullday	0.5608	0.5462	0.5908	0.6054
Fullday.2	**0.5816**	**0.6827**	**0.6533**	**0.6858**

Appendix D10: Performance feature sets with KNN and Logistic Regression

Table D45

Performance KNN by accuracy

Accuracy	*Full*	*Full.2*	*Top20s*	*Top20s.2*	*Top25*	Top25.2	*Fullday*	*Fullday.2*	*Mean*
Baseline	**0.7165**	**0.6417**	**0.7071**	**0.651**	**0.7644**	**0.7194**	**0.7026**	**0.6905**	
Featureset 1	0.5366	0.5561	**0.6143**	**0.6913**	0.6575	**0.7347**	**0.6103**	**0.5833**	0.63
Featureset 2	0.5638	0.5508	0.6071	0.651	**0.6795**	0.7041	0.6077	**0.5833**	0.62
Featureset 3	**0.5662**	**0.5588**	0.5893	0.6242	0.663	0.7296	**0.6103**	0.5714	0.62

Table D46

Performance Logistic Regression by accuracy

Accuracy	*Full*	*Full.2*	*Top20s*	*Top20s.2*	*Top25*	Top25.2	*Fullday*	*Fullday.2*	*Mean*
Baseline	**0.7165**	**0.6417**	**0.7071**	**0.651**	**0.7644**	**0.7194**	**0.7026**	**0.6905**	
Featureset 1	0.6631	0.639	0.625	**0.6913**	0.7041	0.6888	**0.6487**	**0.7143**	0.67
Featureset 2	0.6584	**0.6471**	0.6286	0.6779	0.7096	0.6837	0.6359	0.7083	0.65
Featureset 3	**0.6844**	0.623	**0.6571**	0.6644	**0.7123**	**0.7398**	**0.6487**	0.6964	0.66

Appendix E: Code

Code can be found on https://github.com/xCarien/ThesisMentalTiredness