Classifying Emotional State Based on App Usage Behavior and App Category

Siyou Liu

STUDENT NUMBER: 2019292

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE IN DATA

SCIENCE & SOCIETY

DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE

SCHOOL OF HUMANITIES AND DIGITAL SCIENCES TILBURG
UNIVERSITY

Thesis committee:

Dr. Andrew Hendrickson
Dr. Merel Jung

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
December 2019

Abstract

The goal of this research is to examine what method can be used to transform the Likert scores of different emotions into one emotional state indicator, and how accurately can Random Forest Classifier be used to classify people's emotional state based on their app usage behavior and app category. The research question is*: How accurately can people's negative emotional state be classified by their app usage behavior and app category?* Whilst much previous research investigated the association between people's phone usage and emotion, this research sets out to examine the joint effect of app usage like duration, frequency, earliest usage time. etc., together with six different types of app categories, which provides deeper insights into the app usage behavior. The app usage dataset used in this research was generated by software, which is more reliable compared to self-reported app usage activities manually filled in by the users. Prior to the classification, this research also used a dataset that contains eight different negative scores, measured on a five-point Likert scale, to create the target variable. To be able to classify emotional state instead of discrete emotion, the Likert scores of these eight variables were transformed into one emotional indicator using k-means clustering and principal component analysis, and resampling method as well as feature selection technique based on feature importance was used for further improving the model accuracy. By the end of the research, an accuracy of 90% was achieved.

Keywords: app usage, app category, emotional state, emotion classification, k-means clustering

Introduction

This paper aims to draw on a systematic, extensive research into whether negative emotions can be clustered into multiple groups to reflect people's emotional state, and if the smartphone app usage behavior and the category of apps can be used to classify the emotional state. Classifying emotions is a major area of interest within the field of emotion research. Scholar have long debated whether emotions should be treated as basic, discrete mental states with their own characteristics, or they should be characterized on a dimensional basis in groupings since a grouped emotional indicator can be used to explain the empirical observations in affective neuroscience more adequately using an interconnected neurophysiological system (Posner, Russell & Peterson, 2005).

As with the growth of smartphone usage, its negative impact on emotions has received considerable critical attention from researchers. Several studies examined and revealed the association between people's emotions or mental disorders and their addictive smartphone usage behavior (Billieux et al., 2015; Augner & Hacker, 2011). However, the measurement of smartphone addiction seems to be rather indistinct. Lin et al. (2015) mentioned that although smartphone addiction can be considered as a type of internet addiction, it cannot be measured the same way: the 'traditional' internet addiction behavior can be reflected by a significant degree of time distortion, but smartphone usage is generally much shorter, frequent and fractional. Also, most of the studies collected mobile phone usage data using the form of self-reporting, which can be unreliable as participants may have the tendency to fill in the data heuristically. Vanden Abeele, Beullens and Roe (2013) indicated in their research that while light phone users always overestimate their actual phone usage, heavy phone users tend to underestimate their phone addiction and dependency on their phones. This means that in order to measure the extent of smartphone addiction, more representable measurement metrics are needed.

Furthermore, while people's emotions can be affected by the ways they use the apps, it may as well be affected based on the characteristics of the apps. For instance, although social media apps are divided into various categories based on their content or function, there has been little discussion about the types of social media and whether they would affect people's emotions differently, which makes the generalizability of much published research on this issue problematic.

This research provided a good opportunity to examine whether k-means clustering can be used in combination with PCA for classifying emotional state. Furthermore, it advanced the understanding of the association between phone usage behavior, app category, and emotional state, by demonstrating the statistical evidence generated by modern machine learning supervised techniques. The research question is*: how accurately can people's negative emotional state be classified by their app usage behavior and app category?* The sub-questions are:

● What clustering method can be used on a set of discrete emotion Likert scores for emotional state categorization?

● How accurately can Random Forest classifier be used to classify people's emotional state based on their app usage behavior and app category?

● What is the most situation resampling method to further improve the accuracy of the Random Forest model?

● What is the effect of feature selection based on feature importance on the accuracy of the Random Forest model?

By the end of the research, an emotional state indicator based on the Likert scores of different negative emotions was created, and it was used as the target variable for the Random Forest classifier, while app usage behavior and app category were used as features. The classification model was further improved using resampling techniques and feature selection technique based on feature importance. Based on the results, it can be concluded that a combination of undersampling and oversampling method SMOTEENN improved the model most (by 16%), but the feature selection method based on the feature importance did not work well on improving the accuracy. The final model has an accuracy of 90%.

Related Work

Emotion Classification

Emotion classification has long been a topic of great interest in emotion research and affective science, which consists of different ways of identifying one emotion from another. Scholars have long debated how emotions should be distinguished from each other, but to date, there are two major viewpoints: discrete emotion theories and dimensional models of emotion. Discrete emotion theories, first emerged in the 19th century (Colombetti, 2009), claimed that emotions are discrete constructs which should be treated as separated categories instead of a combined emotional state, since each emotion has unique, particular characteristics which differentiate it from the others (Ekman, 1992; Colombetti, 2009). However, Barrett, Gendron & Huang (2009) offered contradictory findings about the discrete emotion theories, criticized Colombetti's theory and point out that it was not in line with any scientific evidence generated from the neuroscience field, for instance, when Posner et.al (2005) revealed that some emotions which can be treated as a group usually trigger the same area within the brain. Furthermore, grouped emotions also seem to be more effective in terms of emotion research. Nwe, Wei & Silva (2001) conducted a classification using speech-based emotions, where they found that using grouped emotion as labels improved the accuracy by 10% to 23% compared to ungrouped emotion.

Emotions and App Usage Behavior

Mobile phone has become the most widely used mobile device for people’s daily needs for access of information, communication with others and leisure activities, leading the rise of software programs known as applications (apps), which not only can perform various tasks but also is an efficient marketing tool for businesses (Hur et al., 2017). Therefore, it is crucial for companies to understand the patterns and phenomenon reflected behind their customers’ app usage data and adjust their online strategies accordingly. However, while companies are making use of such data and improving their app performances, ethically, they should also be aware of whether their apps are affecting users to develop a problematical phone use behavior like phone addiction, which can lead to several physical health and mental health problems. The research revealed that additive phone usage behavior is likely to lead to mental disorders like sleep disorder or depression, and can jeopardize users' social relationships, especially among the younger generation (Augner & Hacker, 2011). Amongst all kinds of addictive behavior of mobile phones, excessive use is considered to be the most common type and associated closely with the negative outcomes of phone addiction (Billieux et al., 2015). Excessive use of mobile phone apps can be quantified in many ways, for instance, Cheung et al. (2018) proposed a measurement method which measures the app engagement by recording the number of weekly app sessions (“loyalty”) and the number of days with app usage (“regularity”), while Lin et al. (2015) measured the daily use duration and frequency of engaging with mobile phones apps. In this research, the duration and frequency of app usage were recorded and used for the analysis, as the main factors of judging whether a user has excessive usage behavior. However, instead of adding random features, it should also be noted that there is a need for parameters to define the extent of ‘excessive use’. How much time spent on an app is considered to be ‘too much’? How can researchers define someone as a ‘heavy phone user’ or ‘light phone user’ of smartphone apps?

RFM Model for Measuring Phone Addiction

RFM (Recency, Frequency, Monetary) model is one of the most traditional marketing models, proposed by two Dutch professors in an issue of Marketing Science in 1995 (Bult & Wansbeek,1 995). Up until now, it is still widely used by marketers to measure their customers’ brand loyalty and analyze customer value (Qiasi et al., 2012). Birant (2011) described RFM Analysis as a marketing technique used which divide customers into various groups to identify customer value and predict their future purchase behavior, by using three key factors: how recently a customer has purchased (recency), how often the customer purchases (frequency), and how much the customer spends (monetary). By dividing the customers into N groups based on their score for each factor, a segmentation of N*N*N will be generated to define the customer value. For example: if customers are into two groups based on their recency, frequency and monetary value, a segmentation of 8 groups will be generated. The group with the highest frequency, monetary value and the lowest recency value is considered to have the highest customer value, as they purchase more frequently, spend more money and the last purchase date is not long ago (see in Figure 1). By implementing the RFM model, modern data mining techniques can transform easily accessible data into a summary containing a wealth of information of customers((Fader & Hardie, 2009).

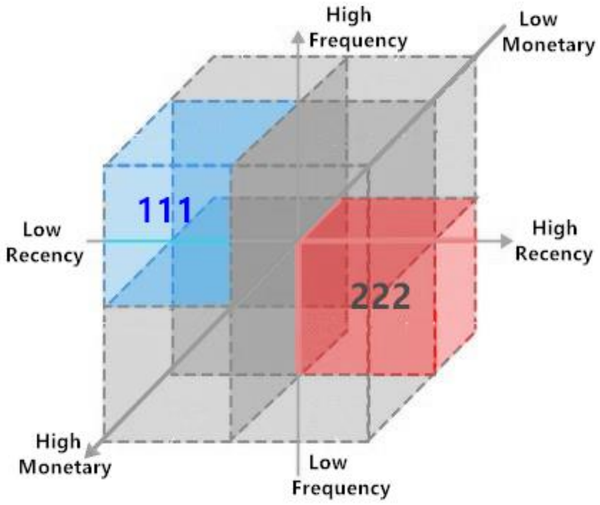


Figure 1. self-designed visualization of the RFM model with two groups for each factor, 8 segments in total. As the most valuable customer group, group 111 has the highest frequency, monetary and lowest recency value; group 222 shows the opposite.

Recent years, the useful mechanism behind RFM model prompted many researchers to start exploring whether its paradigms can be applied in other domains, for instance, Jašek (2014) presented an innovative approach to use website visits and social network interactions.etc as additional data source for RFM modeling to leverage the predictive power of the model and identify the most loyal customers, while Bernabé et al (2015) use the RFM paradigms in their research for quantifying the impact of social media topics and identify popular topics. Hence, in this research, the paradigms of the RFM model were adapted and combined with other predictors to measure the extent of excessive usage. The details of the model will be explained in the method section.

Effect of Different Apps

Up to the second quarter of 2019, there are around 2.46 million and 1.96 million apps for Android and Apple users in the leading app stores (Clement, 2019), which are divided into 35 categories in Google Play ("Android Apps on Google Play," 2019) and 27 categories in Apple store (Apple Inc, n.d.).While different types of apps have different functions or features, it is ambiguous how the characteristics of an app can affect users mentally. Do game apps have the same influence on people like education apps? Researchers have pointed out that while game apps have the tendency to cause pathological behavior, education apps like language learning apps are proven to be effective and easy to use for learning; while using messaging apps during the night can cause sleep disturbances, lifestyle apps designed to track fitness activities or mental health activities enable people to keep up with healthy living habits and have more self-esteem (Augner & Hacker, 2011). Hence, understanding the link between the app category and emotion will help companies to further adjust the features of their apps and prevent users to develop problematical habits of using apps.

This research aimed to examine whether features like app usage have a joint effect on the app category on users' negative emotional state. Unlike the majority of the studies in which the target variable is already pre-defined, this research attempted to build an emotional state indicator using Likert scores of 8 different negative emotions. To achieve this, principal component analysis (PCA) was conducted on these variables, and k-means clustering was applied with both the original data and the reduced components to compare whether PCA was necessary. It is hoped to find results consistent with the scientific evidence from the neuroscience field, that different emotions indeed can be treated as a group to measure the 'emotional state'.

Afterward, the emotional state indicator was as the labels for the classification using Random Forest, with a set of features like usage duration, frequency, app category and other relative features. In order to ensure that the features represent app usage behavior well, a novel approach of adapting the RFM model's paradigm was presented as an effective approach of measuring the extent of the 'loyalty' of users, which in this case reflects the extent of their excessive phone usage behavior. Also, to further improve the accuracy and F1 score of the model, two different methods were used in this research: resampling methods, which is commonly used in machine learning to prevent the potential problem caused by the imbalanced classes of the labels; feature selection method based on the feature importance, to remove features which do not contribute to a higher accuracy of the model. The phone usage dataset used for this research is generated by software instead of manually recorded data, which fills the research gap between the previous studies and gives more insights into the problems addressed in the Introduction section.

Method

To answer the research question, this paper first explored the approach of using clustering for creating an emotional state indicator. Then, while data like app categories and phone usage was used as features, the emotional state indicator was used as the target variable and the ensemble learning technique, Random Forest was used to classify the labels based on the features. Starting with a brief introduction of the raw dataset, this section will then further explain in detail how the modeling approaches for both target variables (emotional state indicator) and the features (app categories and phone usage) were proposed or modified from existed models.

Dataset Description

This research made use of three existing datasets: Phone_Data, App_category and User_Mood, which were provided by researchers from Tilburg University (Hendrickson, Aalbers & Vanden Abeele, under review). Amongst the three datasets, Phone_Data and App_category were used for feature engineering, while User_Mood was used for creating the target variable.

The first dataset for feature engineering, Phone_Data consists of 124 participants' phone usage data recorded using the logging tool MobileDNA, which records the details of each user's phone usage like name of the app used, what apps were used in each session, and starting and ending time of each app usage. In this research, only 4 variables were used: application, user id, start time and end time.

The second dataset App_Category consists of 6 variables, indicating each application's category, the name and their total counts. However, in this research, we are only interested in the categories of the application. There were originally 3 different sets of categories in the App_category, with 59, 70 and 50 different categories, divided according to the functions of the apps and further optimized based on the functions' similarity. These 3 categories were used as a reference to categorize the apps in the Phone_Data dataset, since some of them were overlapping with each other, for instance, category 'Messages','Messaging' and 'Instant Messaging' all contain messaging apps, which should not be differentiated from each other. Hence, the categories will be re-modified as the app category is an important feature to be extracted and prepared for analysis. More detailed information regarding the re-modification of app categories can be found in the Modeling Approaches section below.

The third dataset, User_Mood contains information of 149 participants' self-reported evaluation on their emotions, their daily activities and some descriptive data of the evaluation itself like duration, user ID and date. In this research, we will focus on the negative emotion variables only, which are 8 negative emotion variables, measured using a 5-point Likert scale (for example, 0: not at all anxious, 5: extremely anxious). The imbalanced number of participants in User_Mood and Phone_Use indicated that somehow a few users who filled in the mood survey did not manage to provide their phone usage data, hence, the two datasets were later merged based on their corresponding user ID, after the preprocessing steps and features being extracted and engineered.

Modeling Approaches

Target variable. Based on the literature review, it can be concluded that the 8 negative emotion variables from User_Mood should be transformed into one or multiple grouped variables for a better representation of the emotional state. Most of the research divided emotions into groups based on more subjective data, for instance, Nwe et.al (2001) divided emotions from speech materials based on their similar patterns of energy migration in frequency domain, and Li & Lu (2009) divided the emotions into two groups, happiness and sadness based on the electroencephalography (EEG) signals from the participants' brain. However, in this study, a self-reported Likert scale was used for measuring the emotions. Although the nature of these emotions implied that these variables can be divided into different groups based on the score users given (for instance, 0-2 means positive, 3 means neutral and 4-5 means negative), the fact that the data collected were solely based on users' self-reflection of emotions should be taken into account. Since everyone perceives emotions differently, for example, people who reported themselves as neurotic tend to underestimate while recalling the intensity of negative emotions (Barret, 1997), it may not be scientifically reliable to divide these 8 variables based on the scoring system manually. Michalopoulou & Symeonaki (2017) proposed an automatic approach of using clustering to improve the interpretability of Likert scale raw scores, by applying k-means clustering to the overall scale computed by both summing up and averaging the variables, which were originally on a 1-5 scale. Hence, in this research, we explored whether k-means clustering can be used to cluster the 8 emotion variables into K groups as well, to transform the raw scale scores into a more reliable and simplified indicator with different levels indicating different emotional states.

However, instead of computing the sum and mean of 8 variables, another approach was used to decrease the dimensionality of the variables in this research. The reason for that is the 8 variables are highly correlated, for instance, anxious and stressed are considered to be interchangeable terms and overlap with each other (‘Canadian Mental Health Association’, n.d.). A more in-depth exploratory data analysis was conducted to examine their correlation, based on Figure 2, there is a high correlation between most of the variables, for instance, anxious and gloomy, with a correlation score of 0.7 (see complete correlation matrix in Appendix A). Hence, Principal component analysis (PCA) was conducted on these 8 original variables. PCA can convert a set of possibly correlated variables into a set of values of linearly uncorrelated variables, reducing the dimension of data while still retaining most of the information (Karamizadeh.et al, 2013). This enabled the study to retain most of the information from the 8 emotion variables while eliminating some dimensions to decrease the chances of overfitting and make it easier to visualize the result of k-means clustering as well.

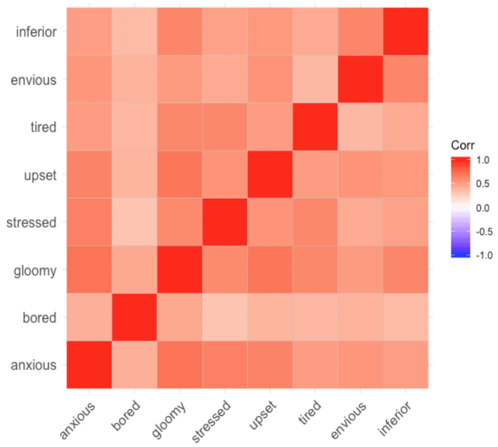


Figure 2, correlation heatmap of each negative emotion variable

Whole research. Below in Figure 3 is the process mapping of the whole research. Starting with a k-means clustering, the raw Likert scores of the 8 emotion variables were transformed into one variable with different levels representing different emotional states of the users (Cluster A). The dimensions of the 8 emotion variables were further reduced using PCA, and the reduced components were used for clustering as well (Clustering B). Then, both clusters before and after PCA were used as the labels to train two random forest classifiers (Model 1 and Model 2), and their results were compared to check whether PCA was necessary for the target variable. Afterward, the model with the highest accuracy and F1 score was selected to be further improved using resampling techniques, and a feature importance list was generated based on the model to explore the features' contribution for the accuracy. The features which did not contribute to the accuracy were removed to check whether the accuracy could be further improved.

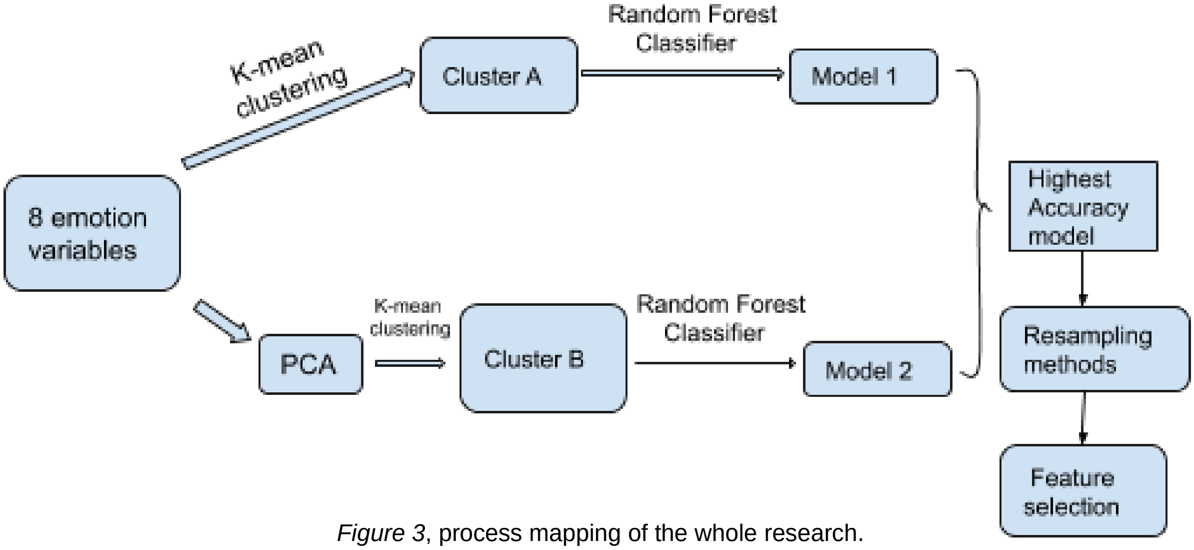


Figure 3, process mapping of the whole research.

Experimental Setup

This section includes three parts: preprocessing, feature engineering and implementation. Both R studio and Python notebook from Google Colab were used for the preprocessing and feature engineering part, but the model training part was conducted only in Google Colab. Starting with the preprocessing process, the feature engineering process of the whole research will be described in detail and the implementation process with the selected algorithm and evaluation will be outlined.

Preprocessing

Phone and app category dataset. Several R libraries, dplyr and lubridate were used to clean and transform the data from the Phone_Use dataset. Firstly, the category of each app was assigned to the Phone_Use dataset based on the information from App_Category. After removing the missing values and duplicated dataset, the cleaned Phone_Use contains 455200 data points, which are the data of unique usage for each app from 124 users in a duration of 34 days. The end and start time of each data point were transformed into Date & Time object and a date was added for each app usage, which is a crucial step for feature engineering later on.

Mood dataset. This study aims to classify people's negative emotional states based on their app usage behavior and different app categories. Hence, the 8 negative emotion variables, Anxious, Bored, Gloomy, Stressed, Tired, Upset, Envious and Inferior in the original User_Mood dataset were used as the core elements to determine the target variable, Emotion_status. The 8 variables have a scale from 0-5, but in this study, they were treated as numerical scores of negative emotions, which means a higher score reflects a more negative emotion.

Prior to the feature engineering, R library dplyr and ggplot2 were used to clean, transform and visualize the User_Mood dataset. First of all, there are a lot of missing data in the dataset due to the fact that some of the surveys were canceled or the session expired before participants actually finish and submit them. Hence, all the rows with missing values were removed. Also, although each emotion was measured using a 5-point Likert scale, there were 4 rows with extreme values for each variable, which were also removed from the dataset (see in Figure 4). The daily sentiment score for each variable was then computed for each user. As mentioned before, since some participants' phone usage data was missing, the User_Mood dataset was matched with the final Phone_Data dataset based on participants' user ID. By the end of preprocessing, the new User_Mood dataset contains 2265 data points, which are the daily mean scores of 8 negative emotions from 120 users in a duration of 34 days.

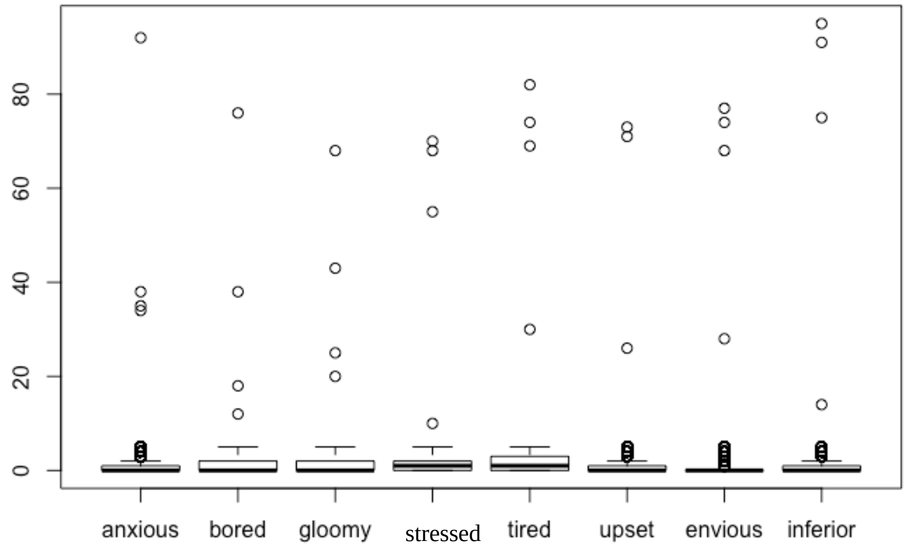


Figure 4, the boxplot of Likert scores of eight negative emotion variables from the origin User_Mood dataset

Feature Engineering

App category features. As mentioned in the Dataset Description, since the categories in the App_category dataset seem to be redundant and not represent well how the apps differ based on their function, they need to be re-modified into bigger categories for a more effective analysis. The category lists from Google Play and Apple Store were mined and compared, together with a phone app category list generated by Böhmer, Hecht, Schöning, Krüger & Bauer in 2011. By looking at their overlapped categories, it can be concluded that Communication, Productivity, News, Entertainment, Social, Education and Utility are common categories that are often included in the category list. Then, the three category lists in the App_category dataset were also investigated and compared to explore how the apps were divided based on their functions and the similarities of their functions. In the end, a category list with 6 different categories was generated, which are: Lifestyle, Social Network, Communication, Utility & Tools, Game & Entertainment, News & Information Outlets. A more detailed category list can be found in the Appendix B. The app category was then assigned to each app, which will be further explained in the preprocessing section below.

RFM & phone usage features. Aside from the app category, most of the phone usage features were also not originally included in the dataset, thus a lot of computation needs to be conducted for feature extraction. In the literature review, the paradigm of the RFM model was clearly explained and used as an inspiration for creating the Recency, Frequency and Monetary features for this research. However, since we want to look at the effect of phone usage on a daily basis, the Recency score needs to be calculated on a daily basis as well. The best way to adapt the Recency calculation formula into this research is to calculate the time difference between midnight 00:00 and the latest time the user ends his app usage, measured in seconds. The adapted formula was created to illustrate the calculation (see Figure 3). For instance, if a user used Facebook until 20:20, the time difference should then be 3 hours and 40 minutes, which means it has a Recency score of 13200. However, by looking at the session data in the original Phone_Use dataset, it should be noted that there are a lot of users still use their phone after midnight, which makes the Recency score not entirely reliable. Thus, another time difference variable, Earliest_time, which measures the time difference between midnight and the earliest app usage was added as a feature. If the user used Facebook at 01:00 at midnight, the Recency score may seem unreliable, but the Earliest_time will reflect that he used this app very early in the morning, which can be also considered as a sign of excessive usage.

$$(\boldsymbol{T}_{midnight} - \boldsymbol{T}_{latest}) = \boldsymbol{T}_{diff}$$

Figure 3, adapted formula to calculate Recency score, the time difference between midnight and latest app usage time

The Frequency and Monetary score in this model are relatively easier to be interpreted. The Frequency would be the total number of users accessing an app on a daily basis, while the Monetary would be the average time per day a user spend on an app. To calculate the Monetary score, the total duration as the numerator in the formula in Figure 4 was extracted from the dataset. Furthermore, in order to gain more insights into the relationship of app categories and engagement behavior and their joint effect on people's negative emotional state, relative features like the proportion of total duration, the proportion of frequency on a daily basis for each category were also calculated.

$$\frac{\mathbf{T}_{duration}}{\mathbf{N}_{frequency}} = Monetary$$

Figure 4, adapted formula to calculate Monetary score (average duration per day)

Other features. a few common descriptive statistics were also added as features: the minimum duration spent on an app category, the maximum duration spent on an app category, the standard deviation and the variance of duration time of each app category on a daily basis.

After adding these features, the reshape2 package from R was used to cast the dataset into a wild format. Each feature was further split by the categories of the app and become 6 new features, except the app category feature itself. For instance, the frequency feature was reshaped into 6 new frequency features, each contains the frequency number of each category. If the user did not use one specific type of app on that day, the frequency value would be missing. In that case, the missing value is replaced with a 0. After reshaping the dataset, the final feature dataset has in total 66 columns, includes the user ID and date.

Lastly, two more time-related features were added: day of the week and day of the month, to measure whether using apps on different days affects people's emotions differently. Prior to the training, all features were standardized using the standardscaler from sklearn preprocessing library.

Implementation

K-means clustering & PCA. This part addressed how the first research question can be answered. After the preprocessing, k-means clustering was conducted to cluster the 8 emotion variables into K groups, using scikit-learn, pandas and numpy library in Python. Since K-mean clustering does not have labels as an unsupervised machine learning technique, the silhouette score was used to evaluate the performance of the clustering. K-means clustering with k value from 2 - 8 was conducted to select the right amount of K, and various graphs were plotted using matplotlib, mpl_toolkits, and plotly.express to evaluate the final clustering manually.

R was used for PCA analysis since the FactoMineR and factoextra library from R was easier to use for extracting and visualizing the results of PCA and help to identify which components should be selected. A common but mostly used approach for selecting components is to first choose the one with the highest explanatory power and shift to the one with the lowest, choosing the components with an aggregated eigenvalues, also called the cumulative percentage of variance, of 80% or 90% (King & Jackson,1999). Hence, the plot of eigenvalue was generated for selecting components. Furthermore, aside from using the traditional approach of looking at the cumulative percentage of variance, FactoMineR also has a lot of metrics that allow researchers to look at data like the quality of the representation, contribution of each component.etc (Lê, Josse & Husson, 2008). Quality of the representation, also called squared cosine, varies from 0 to 1, indicates the importance of a component for a given observation, which enables researchers to check how well-represented these variables are by each component (Abdi & Williams, 2010). Hence, the squared cosine was investigated, and a squared cosine plot was created to illustrate the importance of each component for the variables. The result of PCA, as well as the statistical evidence for components selection, can be found in the Result section.

After the selection, the chosen components were used for conducting a new clustering. The cluster with the best silhouette score (‘Cluster B’ in Figure 3) was used as the final label for the Random Forest Training with PCA. However, to examine whether PCA was really necessary, another clustering using the raw Likert scores (‘Cluster A’ in Figure 3) was also conducted and used as label to train a random forest classifier, after which their results were compared.

Random Forest Classifier. In the last step, the implementation part is hoped to answer the second research question. After the emotion indicator was created, it was used as the final target variable to be trained on the Random Forest classifier. The first fitted model used the data which was split into 20% of test set and 80% training set, with n_estimators = 50 and max_depth = 5. After fitting the model on both labels before and after PCA (‘Model 1’ and ‘model 2’in Figure 3), the results were compared to reveal whether PCA was necessary on the 8 emotion variables. The model with the lower accuracy and F1 score was used as the baseline model. Furthermore, in order to examine whether the model generalizes well on new data, a 5-fold cross-validation was conducted on both models, and the mean and standard deviation of the cross-validated accuracy was used to evaluate whether the model would overfit.

Resampling Methods. To further improve the model and answer the third research question, the problem of class imbalance of the labels was addressed by using different resampling methods to change the number of samples extracted from each label. Two basic resampling methods from Python's imbalanced-learn library were used: RandomUnderSampler for undersampling and BorderlineSMOTE for oversampling. Undersampling method is used to select a number of majority class to match with the number of minority class, while oversampling does exactly the opposite. However, it should be noted that each method has its own disadvantages: while the undersampling method could discard potentially useful data, the oversampling method is very likely to overfit (Weiss, McCarthy & Zabar, 2007). Hence, two other resampling methods SMOTEENN and SMOTETomek were added into the analysis. These two methods combine both oversampling and undersampling techniques, which can prevent losing useful information about the data and decrease the chances of overfitting (Lemaître, Nogueira & Aridas, 2017). The resampling method with the best model performance was selected as the very final model, and its hyperparameters 'n_estimators' and 'max_depth' were tuned using GridSearchCV from sickit-learn library. The cross-validation in the GridSearchCV was also set to 5, to match with the first fitted model by using 20% of the data as the test set, and the 'n_estimators' was set on with a list of [10,20,50,100,300,500], while 'max_depth' was set on a list of [3,7,10,20,30].

Feature Selection. Lastly, to answer question 4, the feature importance of all 68 features was calculated. Random Forest classifier has a default feature importance function which calculates the impurity-based feature importance, however, researchers in the data science field have long debated the disadvantages of it. Pedregosa et al. (2011) explained in their famous scikit-learn user guide that the Random Forest's feature importance is computed using the training set, which means even if some features are not predictive of the target variable, they can still be listed as important features if the model by chance used them to overfit. Hence, in this research, the permutation importance from scikit-learn library eli5 was chosen instead of the default one. Permutation importance from eli5 calculates a feature's importance by looking at the model error when the feature is removed from the feature set using unseen test data, which makes it more reliable. Based on this permutation importance list, we can look into which features were more important to the model and drop the ones which actually do not contribute to the accuracy of the model, by using sckit-learn's feature selection library and finetuning the threshold, which is criteria to remove any features with a weight of contribution below a certain number, to examine whether feature selection can further improve the model.

Result

K-means clustering

K-means clustering before PCA. Figure 6 illustrates the silhouette score result of Cluster A, using a range of 2-8 as the k value on the original 8 emotion variables before PCA was conducted on these variables. Based on the result, it can be concluded that the 8 variables should be clustered into 2 groups, as it has the highest silhouette score, 0.425. Due to the fact that k-means clustering always randomly assigns labels on different runs, to check the meaning of each automatically generated label, a mean plot on the 8 variables grouped by each label was plotted. As shown in Figure 7, class 0 overall has a lower mean compared to class 1, which indicates that all data points assigned to class 0 perceive lower negative emotions, while data points in class 1 reflect that users think they perceived more negative emotions. Since a higher emotion score indicates a higher negative emotion, class 0 contains the users who report their emotions as non-negative, while class 1 means negative.

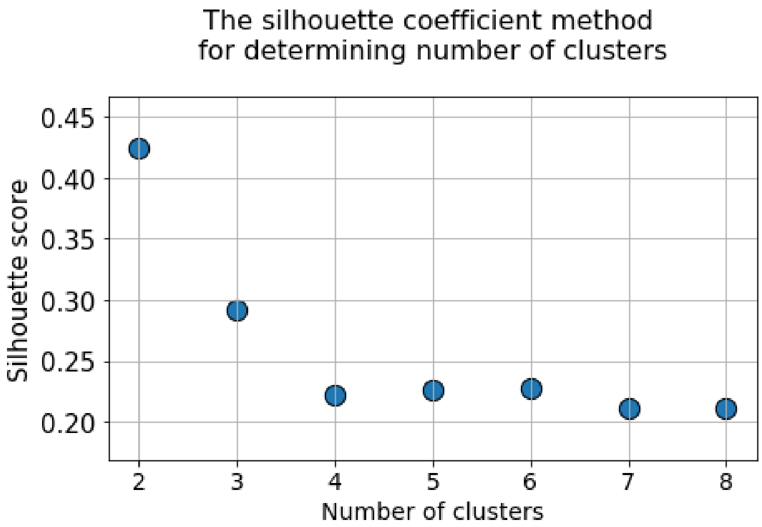


Figure 6, silhouette scores of K-means clustering, with a K value of 2 - 8

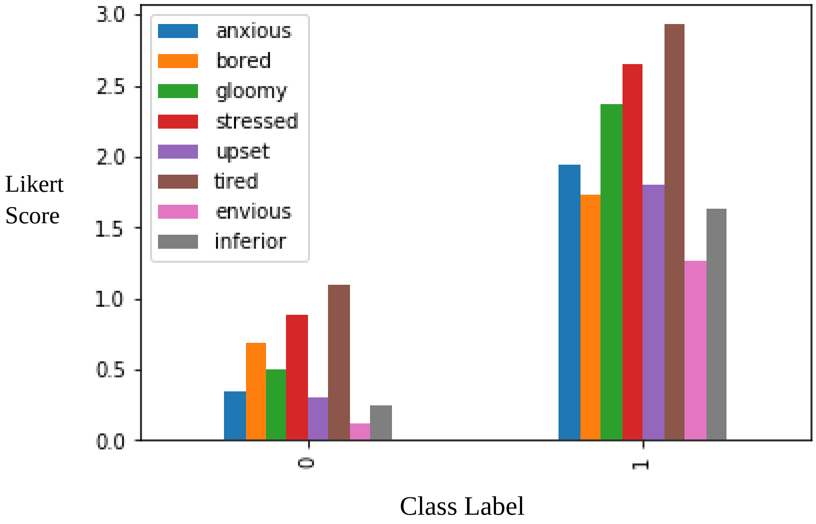


Figure 7, mean plot of each variable grouped by k-means clustering

Principal component analysis. PCA was further conducted on the 8 original variables and in total 8 components were created. Based on King & Jackson (1999)'s theory mentioned in the Experimental Setup section, it can be concluded that in this research, at least the first 4 components should be selected, as they account for 83.1% of the total variance (see in Figure 8).

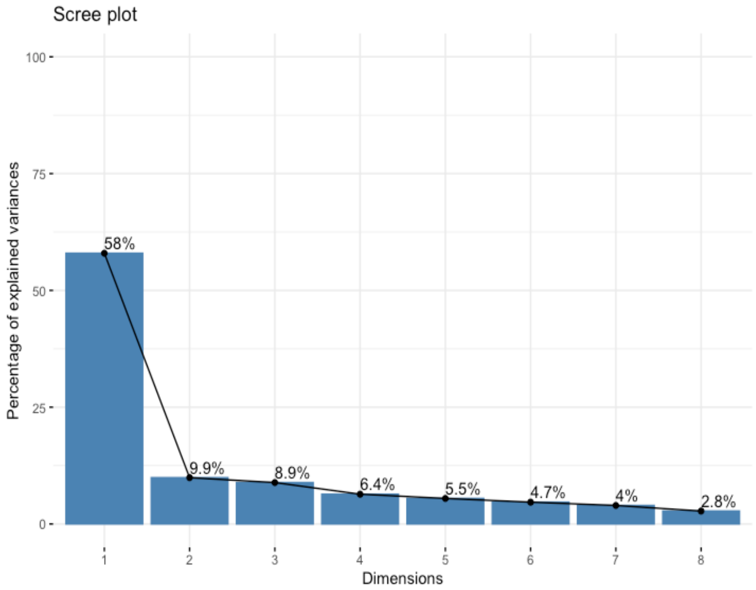


Figure 8, Scree plot of each component's percentage of variance, generated by the factoextra library in R

By computing the squared cosine of the first 5 components (see in Figure 9), a better overview of the contribution of these components to variables were illustrated. The first component contributes to almost every variable, which makes it the most important component. The second one is important for variable Bored, Stressed and Envious, and the third variable again contributes to Bored, Envious and Inferior. In total, the three components contribute to an average of 76.7 % of all the variables, with a minimum of 66.3% to Upset and a maximum of 98.4% to Bored. The 4th and 5th component, however, does not represent the variables well, especially the fifth component, which hardly has a strong contribution of any variable. The fourth component, in spite of its small contribution to other variables, it still seems to contribute to 15.1 % to variable Tired and 13.8% to Anxious (see in Appendix C). Hence, the final decision was to select the first 3 and the first 4 components, with a cumulative percentage of the variance of 76.7 % and 83.1 % to conduct two K-means clustering separately.

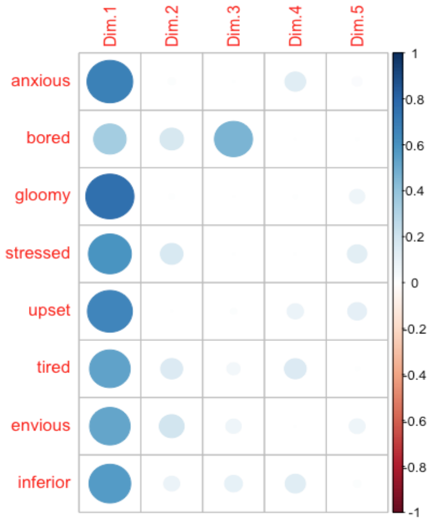


Figure 9, the plot of each variable’s contribution of the top 5 components

K-means clustering after PCA. Two k-mean clustering using a range of 2-8 as the k value were conducted on the first 3 and first 4 principal components, and the results are illustrated in Figure 10 below. Both PCA improved the cluster quality compared to before, but the cluster with 3 components has the best silhouette score of 0.5, while the 4-component one has only 0.47. Hence, it was decided to use the first 3 components to create a new clustering, which is Cluster B, for the Random Forest classification. Figure 11 shows that just like the labels from Cluster A in Figure 7, class 0 and 1 for Cluster B represent people who are in a non-negative and negative emotional state.

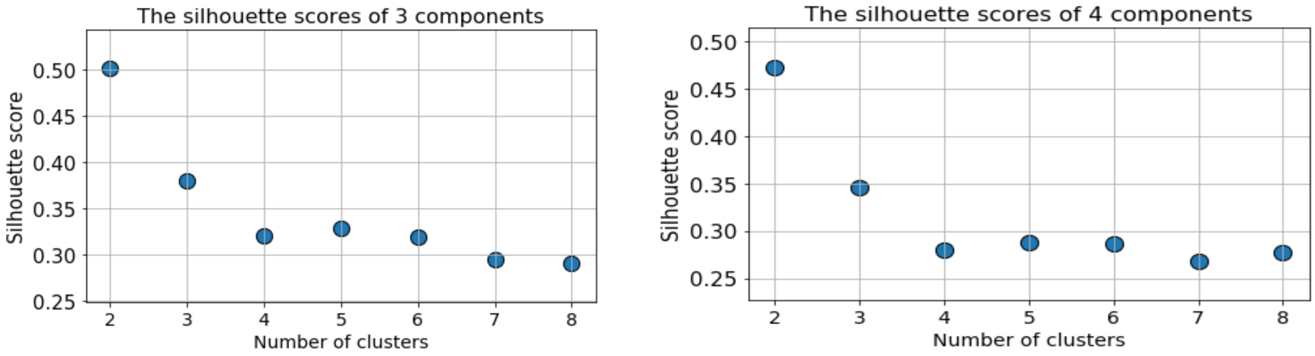


Figure 10, silhouette scores of k-means clustering with 3 components and 4 components, with a K value of 2 – 8

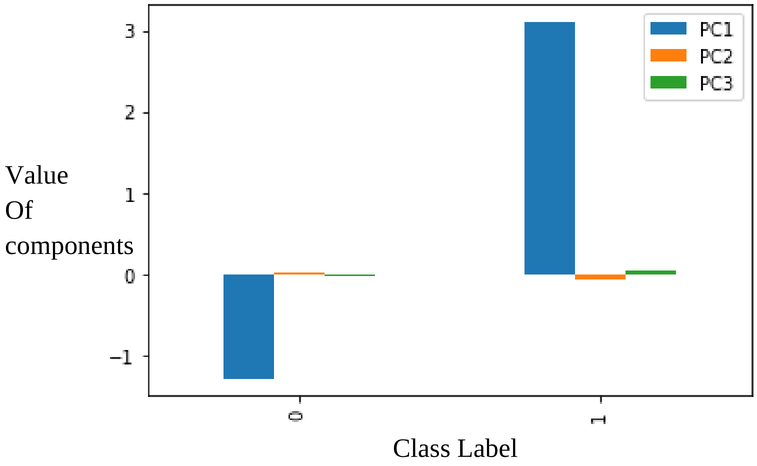


Figure 11, mean plot of the first 3 components grouped by k-means clustering

Overall, the silhouette score of Cluster B was higher than Cluster A. As the dimensions were reduced to 3 components in Cluster B, it was also easier to visualize the quality of the clustering. As can be seen from Figure 12 below, the data points from two classes of Cluster B hardly overlap with each other, indicating a good quality for the cluster (see Appendix D for the multi-angle 3D plot generated by plotly.express library). K-means clustering in combination with PCA managed to transform the 8 groups of Likert scale raw scores into one variable which retained most of the information while reflecting the positive/negative emotional state of users.

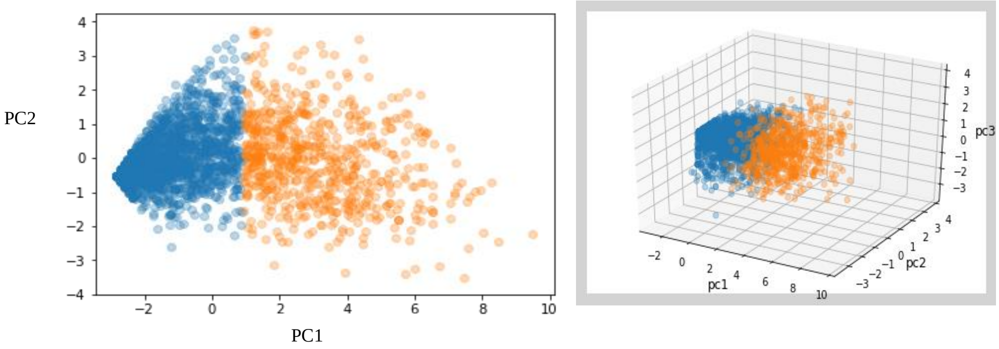


Figure 12, 2D and 3D visualization of Cluster B based on the first 3 components

Random Forest Classifier

Below in Table 1 is the classification report of test set trained with random forest classifier, using Cluster A and Cluster B separately as a target variable. Both models used 20% of data as the test set and 80% as the training set, with hyperparameters n_estimators = 50 and max_depth = 5. As we can see, Model 2 has a slightly higher accuracy, but both models do not differ much. After conducting the 5-fold cross-validation, the mean and standard deviation of the cross-validated accuracy of Model 1 was 0.711 and 0.015, the mean and standard deviation of the cross-validated accuracy of Model 2 was 0.716 and 0.015. This means both models did not show a sign of overfitting since they all have low variance. The mean accuracy did not differ a lot from each other. Due to the relatively lower accuracy score, model 1 is selected as the baseline model for this research.

Table 1

Classification report of the RF model fitted with Cluster A and Cluster B as labels, with 0 being the non-negative class and 1 being the negative class

	Model 1 (label:Cluster A)		Model 2 (label:Cluster B)	
Metrics	Class 0	Class 1	Class 0	Class 1
Accuracy	0.71		0.73	
Precision	0.70	0.82	0.72	0.93
Recall	0.99	0.12	1.00	0.10
F1-score	0.82	0.21	0.84	0.19

Resampling methods. Although the accuracy is not extremely low, in both models, the recall score of class 1 (negative class) is not very high, causing a low F1- score as well. This is due to the imbalanced classes clustered by k-means clustering: class 1 has much less label than class 0. Hence, resampling methods were used to fix this problem. The results of four resampling methods after tuning the hyperparameters using grid search are presented in Table 2. It is apparent that the undersampling method did not perform better than the oversampling, in fact, it even did not outperform the baseline model, which is possibly due to the loss of information when cutting off the samples. Surprisingly, the oversampling method did very well, even better than the SMOTETomek method, in spite of the fact that it usually tends to overfit. SMOTEENN method outperformed all other methods, with an accuracy of 0.90, which was selected as the final model to be further improved on.

Table 2

Classification report of the fitted model after different resampling methods, with 0 being the non-negative class and 1 being the negative class

	RandomUnderSampler		BorderlineSMOTE		SMOTEENN		SMOTETomek	
Metrics	Class 0	Class 1	Class 0	Class 1	Class 0	Class 1	Class 0	Class 1
Accuracy	0.65		0.87		0 .90		0 .83	
Precision	0.64	0.66	0.85	0.90	0.99	0 .87	0.80	0 .86
Recall	0.69	0.61	0.89	0.86	0.72	1 .00	0.86	0 .86
F1-score	0.67	0.63	0.87	0.88	0.83	0 .93	0.83	0 .83

Feature selection based on feature importance. The permutation importance list on 68 features of the final model are included in the Appendix E. There were 42 features which showed a positive contribution to the model accuracy, 10 of which were from the Communication category, 8 of them were from Lifestyle category and Game & Entertainment category, 5 from Social Network and 4 from News & Information outlet and Utility & Tools. As for the app usage behavior, 20 features which were related to Duration contributed to the model positively, which means Duration is an important factor in terms of emotion classification. Frequency and its relative features were also contributing to the model accuracy, with 8 features having weight more than 0. However, Recency and Monetary both did not contribute a lot to the accuracy, so as Earliest_time.

Later, the 26 features which had a negative weight in the permutation list were dropped from the feature column using the feature selection method. In order to further improve the model, the final model which was already improved by resampling method was used for testing the feature selection method, instead of just comparing them. Grid search which the same hyperparameters as the resampling method was applied. Furthermore, the threshold was adjusted to a higher amount, which means the algorithm removed more features, including those which actually contribute to the model accuracy, but only in the slightest sense. Table 3 below shows the result of different thresholds, and there was no improvement observed in terms of accuracy, Precision and F1 score.

Table 3

Classification report of the fitted model after feature selection technique

Threshold	0		0.0005		0.001		0.002	
Metrics	Class 0	Class 1	Class 0	Class 1	Class 0	Class 1	Class 0	Class 1
Accuracy	0.90		0.90		0.90		0 .89	
Precision	0.95	0.88	0.95	0.88	0.97	0.87	0.93	0 .87
Recall	0.75	0.98	0.74	0.98	0.72	0.99	0.72	0 .97
F1-score	0.84	0.93	0.83	0.93	0.83	0.93	0.81	0 .92

Discussion

The goal of this research is to examine whether the Likert scores of different emotion variables can be transformed into one emotional state indicator. Furthermore, it was also expected to find whether app usage and app category can be used to classify people's emotional state using Random Forest classifier, and how the model can be further improved using resampling and feature selection techniques. By conducting a k-means clustering in combination with principal component analysis, this paper managed to build an emotional state indicator which categorized the Likert scores of the 8 original emotion variables into two classes (0: non-negative, 1: negative). This result was in line with the previous findings from other research on emotion classification, that different emotions which are highly correlated can be treated as a group. Also, this research confirmed the feasibility of the automatic approach of using clustering to improve the interpretability of raw Likert scores (Michalopoulou & Symeonaki, 2017), which also answered the first research question from the Introduction section.

Although conducting a PCA on the 8 variables in combination with k-means clustering did not drastically improve the Random Forest model accuracy compared to the clustering without PCA (73% and 71%), the silhouette score of both models did prove that doing a PCA would result in a better clustering. In fact, the minor difference between both Random Forest models reflects the fact that PCA was able to retain most of the information from the original variables, which result in similar accuracy value.

Overall, the resampling methods, RandomUnderSampler, BorderlineSMOTE, SMOTEENN and SMOTETomek, all fixed the problem of class imbalance, which initially caused a low Accuracy and F1-score. The hyperparameters were tuned using GridSearchCV, which also improved the accuracy of in combination with each resampling method, and the final model achieved 90% accuracy by using SMOTEENN.

The permutation importance list revealed a lot of interesting facts about the features and their relationship in terms of classifying people's emotional state, for instance, Communication and Lifestyle category are very important app categories for the model accuracy, as well as features which are measuring Duration and Frequency. This is in line with the literatures presented in the Related Work section, that excessive use like using the smartphone for a long time or very frequently and using messaging apps or lifestyle apps indeed affect people's emotional state. Although it was highlighted in the literature review that Monetary is also an interesting factor to measure addiction level of smartphone users, in this research, Monetary and Earliest_time both did not contribute much to the model accuracy, which means the time of the day when the users using the apps does not affect their emotional state. What was the most surprising result was when using feature selection technique to drop the features which did not contribute to the model accuracy, no improvement was observed from the statistical evidence. In fact, when removing more and more features, the accuracy became slightly lower and lower each time. This is on contrary with what has been discussed in machine learning field, that dropping 'harmful' features would result in a better accuracy. This is probably due to the fact that the threshold value of the feature selection algorithm was not the best value, that during the analysis too many or too less features were removed. Overall, this paper managed to apply a k-means clustering for improving interpretability of raw Likert score to emotion classification and proved that PCA is effective in terms of improving the result of clustering, especially when visualization of the clustering is needed. On some degree, this research also revealed the relationship between app usage, app category and emotional state, though more exploration data analysis should be conducted to check the direction of the relationship. It also proved that resampling method can be a useful tool for fixing class imbalance problem in classification analysis, but it is important to choose the correct method based on the dataset.

For future research, it is recommended to further finetune the threshold of feature selection library and investigate why feature selection method did not have a significant positive effect on model accuracy. It would also be interesting to use other classifiers, for instance, Support Vector Machine and Neural Network, to examine whether they can perform better than Random Forest in terms of classification.

Acknowledgments

I would like to express my gratitude to my supervisor, Dr. Andrew Hendrickson, who granted access to this dataset and provide me useful comments, remarks and encouragement through the learning process of my research proposal and the final thesis. Furthermore, I want to give my thanks to Dr. Giovanni Cassani, who also evaluated my research proposal, and was involved in the first couple of thesis meetings to provide me advice and insights.

I would also like to acknowledge Dr. Merel Jung as the second reader of this thesis, for taking her time evaluating and giving me valuable feedback for improving my first thesis draft.

References

Android Apps on Google Play. (2019). Android Apps on Google Play. Retrieved from https://play.google.com/store/apps.

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433–459. doi: 10.1002/wics.101

Apple Inc. (n.d.). Categories and Discoverability - App Store. Retrieved from https://developer.apple.com/app-store/categories/.

Augner, C., & Hacker, G. W. (2011). Associations between problematic mobile phone use and psychological parameters in young adults. International Journal of Public Health, 57(2), 437– 441. doi: 10.1007/s00038-011-0234-z

Barrett, L. F. (1997). The Relationships among Momentary Emotion Experiences, Personality Descriptions, and Retrospective Ratings of Emotion. *Personality and Social Psychology Bulletin*, *23*(10), 1100–1110. doi: 10.1177/01461672972310010

Barrett, L. F., Gendron, M., & Huang, Y. M. (2009). Do discrete emotions exist? *Philosophical Psychology*, *22*(4), 427–437. doi: 10.1080/09515080903153634

Bernabé-Moreno, J., Tejeda-Lorente, A., Porcel, C., & Herrera-Viedma, E. (2015). A new model to quantify the impact of a topic in a location over time with Social Media. *Expert Systems with Applications*, *42*(7), 3381–3395. doi: 10.1016/j.eswa.2014.11.067

Bianchi, A., & Phillips, J. G. (2005). Psychological Predictors of Problem Mobile Phone Use. CyberPsychology & Behavior, 8(1), 39–51. doi: 10.1089/cpb.2005.8.39

Billieux, J., Maurage, P., Lopez-Fernandez, O., Kuss, D. J., & Griffiths, M. D. (2015). Can Disordered Mobile Phone Use Be Considered a Behavioral Addiction? An Update on

Current Evidence and a Comprehensive Model for Future Research. Current Addiction Reports, 2(2), 156–162. doi: 10.1007/s40429-015-0054-y

Böhmer, M., Hecht, B., Schöning, J., Krüger, A., & Bauer, G. (2011). Falling asleep with Angry Birds, Facebook and Kindle. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI 11*. doi: 10.1145/2037373.2037383

Bult, J. R., & Wansbeek, T. (1995). Optimal Selection for Direct Mail. *Marketing Science*, *14*(4), 378–394. doi: 10.1287/mksc.14.4.378

Canadian Mental Health Association. (n.d.). What's the difference between anxiety and stress? Retrieved from https://www.heretohelp.bc.ca/q-and-a/whats-the-difference-between-anxiety- and-stress.

Cheung, K., Ling, W., Karr, C. J., Weingardt, K., Schueller, S. M., & Mohr, D. C. (2018). Evaluation of a recommender app for apps for the treatment of depression and anxiety: an analysis of longitudinal user engagement. Journal of the American Medical Informatics Association, 25(8), 955–962. doi: 10.1093/jamia/ocy02

Clement, J. (2019, August 6). App stores: number of apps in leading app stores 2019. Retrieved from https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/.

Colombetti, G. (2009). From affect programs to dynamical discrete emotions. *Philosophical Psychology*, *22*(4), 407–425. doi: 10.1080/09515080903153600

Do, T.-M.-T., & Gatica-Perez, D. (2010). By their apps you shall understand them. Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia - MUM 10. doi: 10.1145/1899475.1899502

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3-4), 169–200. doi: 10.1080/02699939208411068

Fader, P. S., & Hardie, B. G. (2009). Probability Models for Customer-Base Analysis. *Journal of Interactive Marketing*, *23*(1), 61–69. doi: 10.1016/j.intmar.2008.11.003

Hung, G. C.-L., Yang, P.-C., Chang, C.-C., Chiang, J.-H., & Chen, Y.-Y. (2016). Predicting Negative Emotions Based on Mobile Phone Usage Patterns: An Exploratory Study. JMIR Research Protocols, 5(3). doi: 10.2196/resprot.555

Hur, H. J., Lee, H. K., & Choo, H. J. (2017). Understanding usage intention in innovative mobile app service: Comparison between millennial and mature consumers. Computers in Human Behavior, 73, 353–361. doi: 10.1016/j.chb.2017.03.051

Jašek,P. (2014). Analyzing user activity based on RFM models complemented with website visits and social network interactions. In D.Petr, C, Gerhard & O, Vaclav(Eds.), *22nd Interdisciplinary Information Management Talks* (pp.181-189). Podebrady, Czech Republic: Trauner Verlag universität.Retrieved from https://www.researchgate.net/publication/288143147_Analyzing_user_activity_based_on_RFM_models_complemented_with_website_visits_and_social_network_interactions

Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., & Hooman, A. (2013). An Overview of Principal Component Analysis. Journal of Signal and Information Processing, 04(03), 173–175. doi: 10.4236/jsip.2013.43b031

King, J. R., & Jackson, D. A. (1999). Variable selection in large environmental data sets using principal components analysis. *Environmetrics*, *10*(1), 67–77. doi: 10.1002

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: AnRPackage for Multivariate Analysis. Journal of Statistical Software, 25(1). doi: 10.18637/jss.v025.i01

Lemaître,G., Nogueira,F., & Aridas, C. K.(2017). Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning, The Journal of Machine Learning Research, v.18 n.1, p.559-563, January 2017

Lin, Y.-H., Lin, Y.-C., Lee, Y.-H., Lin, P.-H., Lin, S.-H., Chang, L.-R., & Kuo, T. B. (2015). Time distortion associated with smartphone addiction: Identifying smartphone addiction via a mobile application (App). Journal of Psychiatric Research, 65, 139–145. doi: 10.1016/j.jpsychires.2015.04.003

Michalopoulou, C., & Symeonaki, M. (2017). Improving Likert Scale Raw Scores Interpretability with K-means Clustering. *Bulletin of Sociological Methodology/Bulletin De Méthodologie Sociologique*, *135*(1), 101–109. doi: 10.1177/0759106317710863

Nwe, T. L., Wei, F. S., & Silva, L. D. (2001). Speech based emotion classification. *Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology. TENCON 2001 (Cat. No.01CH37239)*. doi: 10.1109/tencon.2001.949600

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*, 2825–2830. doi: arXiv:1201.0490

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and Psychopathology, 17(03). doi: 10.1017/s0954579405050340

Qiasi, R., Baqeri-Dehnavi, M., Minaei-Bidgoli, B., & Amooee, G. (2012). Developing A Model For Measuring Customers Loyalty And Value With Rfm Technique And Clustering

Algorithms. Journal of Mathematics and Computer Science, 04(02), 172–181. doi: 10.22436/jmcs.04.02.07

Weiss, Gary M., McCarthy, K., & Zabar, B. "Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs?." DMIN 7 (2007): 35-41.Retrieved from https://storm.cis.fordham.edu/gweiss/papers/dmin07-weiss.pdf

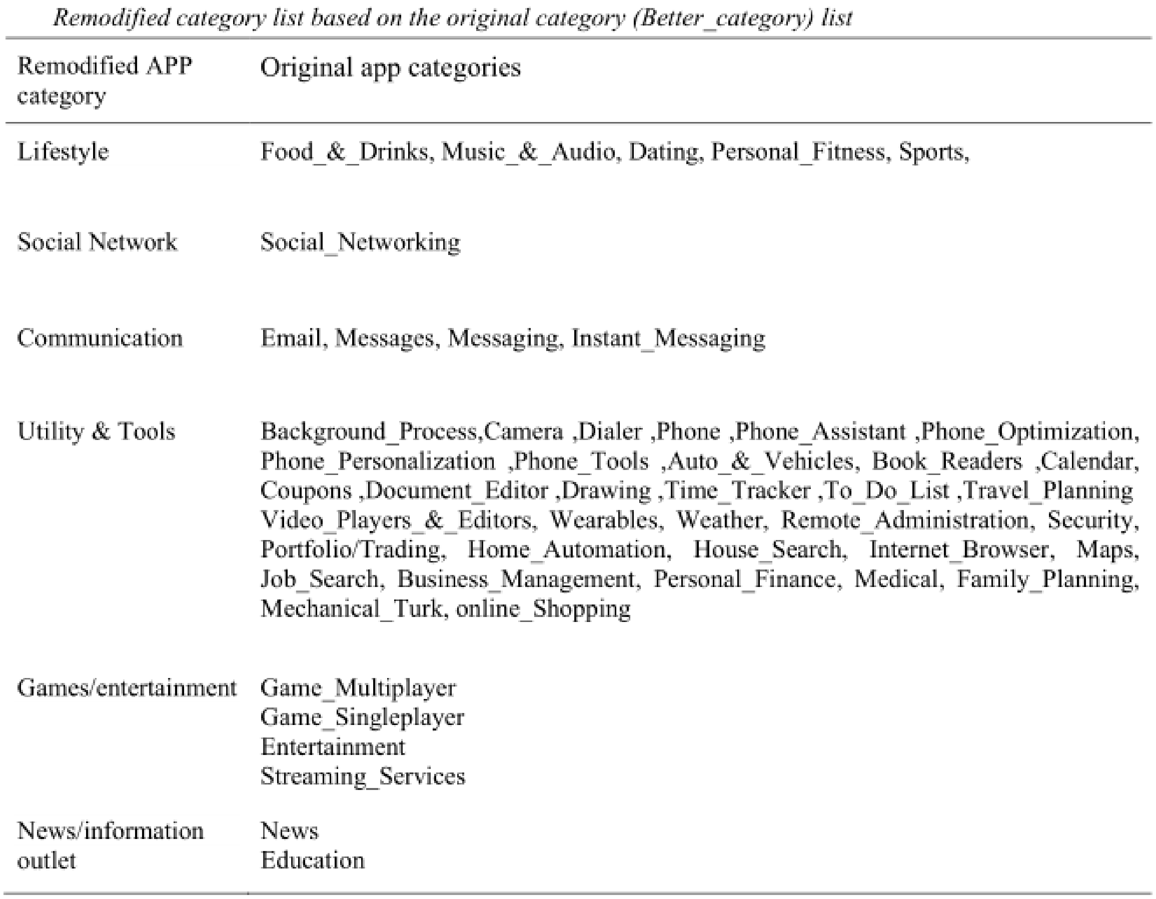
Wu, Q. (2015). Designing a smartphone app to teach English (L2) vocabulary. Computers & Education, 85, 170–179. doi: 10.1016/j.compedu.2015.02.013

Appendix A

Correlation score of each emotion variable

	anxious	**bored**	**gloomy**	**stressed**	**upset**	**tired**	**envious**	**inferior**
anxious	1.0000000	0.4102613	0.6989400	0.6431331	0.6310680	0.5097013	0.5426719	0.4955680
bored	0.4102613	1.0000000	0.4467578	0.3130328	0.3907841	0.3801052	0.4047434	0.3581348
gloomy	0.6989400	0.4467578	1.0000000	0.6046411	0.6911951	0.6115023	0.5184349	0.6230223
stressed	0.6431331	0.3130328	0.6046411	1.0000000	0.5504873	0.6146632	0.4362074	0.4771246
upset	0.6310680	0.3907841	0.6911951	0.5504873	1.0000000	0.5079376	0.5458883	0.5330635
tired	0.509701	0.3801052	0.6115023	0.6146632	0.5079376	1.0000000	0.3705515	0.4396734
envious	0.5426719	0.4047434	0.5184349	0.4362074	0.5458883	0.3705515	1.0000000	0.6214635
inferior	0.4955680	0.3581348	0.6230223	0.4771246	0.5330635	0.4396734	0.6214635	1.0000000

Appendix B



Remodified category list based on the original category (Better_category) list

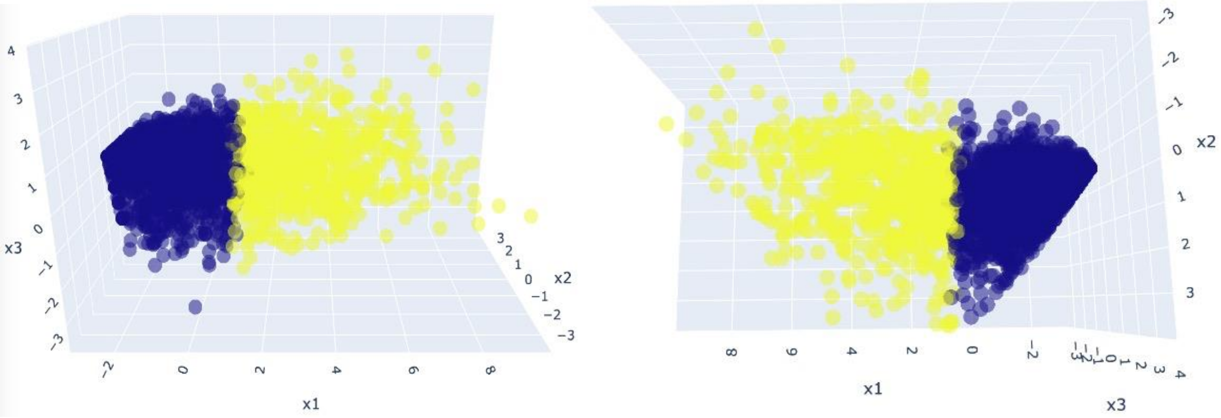
Remodified APP category	Original app categories
Lifestyle	Food_&_Drinks, Music_&_Audio, Dating, Personal_Fitness, Sports,
Social Network	Social_Networking
Communication	Email, Messages, Messaging, Instant_Messaging
Utility & Tools	Background_Process,Camera ,Dialer ,Phone ,Phone_Assistant ,Phone_Optimization, Phone_Personalization ,Phone_Tools ,Auto_&_Vehicles, Book_Readers ,Calendar, Coupons ,Document_Editor ,Drawing ,Time_Tracker ,To_Do_List ,Travel_Planning Video_Players_&_Editors, Wearables, Weather, Remote_Administration, Security, Portfolio/Trading, Home_Automation, House_Search, Internet_Browser, Maps, Job_Search, Business_Management, Personal_Finance, Medical, Family_Planning, Mechanical_Turk, online_Shopping
Games/entertainment	Game_Multiplayer Game_Singleplayer Entertainment Streaming_Services
News/information outlet	News Education

Appendix C

Square cosine of each component with observations

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
anxious	0.6753875	0.0135649272	9.397975e-04	0.1380340482	0.028868927
bored	0.3403236	0.1766259148	4.675037e-01	0.0001513552	0.001949703
gloomy	0.7503098	0.0073146361	4.515064e-05	0.0031108376	0.073917746
stressed	0.5941996	0.1628532086	6.549161e-04	0.0006702349	0.122134785
tired	0.6523565	0.0002580997	1.028296e-02	0.0851249737	0.112627917
upset	0.5325016	0.1549852110	5.461365e-02	0.1509301852	0.003339725
envious	0.5289645	0.1973529544	7.457629e-02	0.0008505022	0.079831547
inferior	0.5623799	0.0805651976	1.008494e-01	0.1313692254	0.014902603

Appendix D



Multiple-angle 3D graphs of the k-means clustering result

Appendix E

Permutation Importance by eli5 library

Weight	Feature	Name
0.0145 ± 0.0096	x44	max_duration_communication
0.0117 ± 0.0055	x1	Day of month
0.0100 ± 0.0076	x25	freq_pro_utility & tools
0.0089 ± 0.0129	x14	dur_pro_communication
0.0084 ± 0.0093	x6	frequency_social network
0.0084 ± 0.0050	x16	dur_pro_lifestyle
0.0072 ± 0.0045	x0	Day of week

0.0072 ± 0.0067	x56	min_duration_communication
0.0061 ± 0.0055	x61	min_duration_utility & tools
0.0050 ± 0.0022	x63	var_duration_games & entertainment
0.0050 ± 0.0042	x57	min_duration_games & entertainment
0.0045 ± 0.0057	x34	monetary_lifestyle
0.0045 ± 0.0027	x46	max_duration_lifestyle
0.0045 ± 0.0125	x42	earliest_time_social network
0.0045 ± 0.0045	x40	earliest_time_games & entertainment
0.0039 ± 0.0083	x58	min_duration_lifestyle
0.0039 ± 0.0076	x8	cat_duration_communication
0.0039 ± 0.0057	x28	recency_lifestyle
0.0033 ± 0.0089	x50	std_duration_communication
0.0033 ± 0.0124	x2	frequency_communication
0.0028 ± 0.0061	x31	recency_utility & tools
0.0028 ± 0.0035	x4	frequency_lifestyle
0.0028 ± 0.0035	x53	std_duration_news & information outlet
0.0028 ± 0.0061	x20	freq_pro_communication
0.0028 ± 0.0035	x37	monetary_utility & tools
0.0022 ± 0.0055	x62	var_duration_communication
0.0022 ± 0.0089	x3	frequency_games & entertainment
0.0022 ± 0.0065	x38	earliest_time_communication
0.0022 ± 0.0022	x52	std_duration_lifestyle
0.0017 ± 0.0067	x59	min_duration_news & information outlet
0.0017 ± 0.0045	x32	monetary_communication
0.0017 ± 0.0057	x36	monetary_social network
0.0011 ± 0.0076	x10	cat_duration_lifestyle
0.0011 ± 0.0057	x9	cat_duration_games & entertainment
0.0011 ± 0.0083	x65	var_duration_news & information outlet
0.0011 ± 0.0083	x30	recency_social network
0.0006 ± 0.0074	x12	cat_duration_social network
0.0006 ± 0.0074	x11	cat_duration_news & information outlet
0.0006 ± 0.0022	x22	freq_pro_lifestyle
0.0006 ± 0.0022	x33	monetary_games & entertainment
0.0006 ± 0.0022	x39	earliest_time_games & entertainment
0.0006 ± 0.0042	x45	max_duration_games & entertainment
0 ± 0.0000	x64	var_duration_lifestyle
-0.0000 ± 0.0079	x18	dur_pro_social network
-0.0000 ± 0.0035	x41	earliest_time_news & information outlet
-0.0000 ± 0.0050	x35	monetary_news & information outlet
-0.0006 ± 0.0022	x29	recency_news & information outlet
-0.0006 ± 0.0042	x5	frequency_news & information outlet
-0.0006 ± 0.0042	x60	min_duration_social network
-0.0011 ± 0.0057	x51	std_duration_games & entertainment
-0.0011 ± 0.0057	x47	max_duration_news & information outlet
-0.0011 ± 0.0045	x23	freq_pro_news & information outlet
-0.0011 ± 0.0045	x26	recency_communication
-0.0017 ± 0.0045	x15	dur_pro_games & entertainment
-0.0017 ± 0.0091	x7	frequency_utility & tools
-0.0017 ± 0.0057	x48	max_duration_social network
-0.0017 ± 0.0115	x54	std_duration_social network
-0.0017 ± 0.0083	x24	freq_pro_social network
-0.0022 ± 0.0082	x66	var_duration_social network
-0.0022 ± 0.0022	x55	std_duration_utility & tools
-0.0028 ± 0.0093	x43	earliest_time_utility & tools
-0.0033 ± 0.0055	x19	dur_pro_utility & tools
-0.0033 ± 0.0042	x17	dur_pro_news & information outlet
-0.0039 ± 0.0045	x67	var_duration_utility & tools
-0.0050 ± 0.0065	x13	cat_duration_utility & tools
-0.0056 ± 0.0070	x27	recency_games & entertainment
-0.0061 ± 0.0065	x21	freq_pro_games & entertainment
-0.0072 ± 0.0057	x49	max_duration_utility & tools