Is simple better? The cost – benefit analysis of models used in predicting procrastination

Student details

Name: R. Ramachandran
Student number: U 710207

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

Supervisor: Dr. A. T. Hendrickson
Second reader: Dr. C. Güven

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

May, 2021

Contents

Abstract

Procrastination has a growing effect in the society. Previous studies have shown the costs of procrastination as a behavioral trait and benefits of using procrastination as self-regulatory strategy. This study aims to find how far variables from different data collection methods could predict trait and momentary procrastination. The two methods of data collection assessed in this study are the one-shot questionnaire method called “Survey Method” measuring trait behaviors and the other is “Experience Sampling Method” assessing behavior repeatedly over time. Four machine learning algorithms are used to build the best predictive model. These are Linear Regression, K-Nearest Neighbors, Random Forest and XGB Regressor. The analyses have showed that combing trait behavior features with momentary behavior traits using XGB Regressor predicts momentary procrastination 23% better than the baseline model. This study highlights the costs and benefits of using these models.

Keywords: Procrastination, Experience Sampling Method, Machine Learning

1. Introduction

1.1. Context

According to Occam's Razor principle, given two theories that make the same prediction, the simplest explanation is preferred to the complex one. Could this hold true for all problems? Of course not, there are always exceptions to rules. But does it hold true to procrastination ? Procrastination has been a widely studied topic due to its complicated nature. Several studies have pointed out the ill effects of procrastination (Tice and Baumeister, 1997) and several more have pointed it out as a self-regulatory strategy to combat burnout and fatigue (Chauhan et al., 2020). Understanding procrastinatory behavior and predicting it effectively would be beneficial in various scenarios.

Procrastination has been seen as a behavioral trait and has been extensively studied to understand its dilatory behavioral tendencies. Researchers such as Chyi-How Lay (1990, 1992, 1993, 1995, 1997), Joseph R Ferrari (1992, 1995, 2005, 2007) and others have termed procrastination as a behavioral trait and studied its continuous ill effects. All these papers have a common thread that participants were asked to report their average procrastination levels. They record their responses through filling a standardized questionnaire. This method of data collection is commonly known as Survey Method. On the other hand, with a technological development, procrastination has also been studied in its momentary manifestations (Wessel et al., 2019, 2020; Aalbers et al., 2021). This has been possible by studying individuals over a certain period in their natural environment. This is known as Experience Sampling Method (ESM).

Here we have a dataset collected from studying smartphone use patterns of students from Tilburg University (Aalbers et al., 2021). The participants had filled in the onboarding survey measuring their personality, perceived stress, fatigue, procrastination levels, etc. The onboarding survey collected information on their individual's trait characteristics. After that, their phone use was logged using the MobileDNA app and through ESM their momentary procrastination, stress, fatigue, and happiness levels were measured using self-reporting surveys over a period of one month.

The goal of this study is to see if data collected during the intake survey performs better when trying to predict procrastination using machine learning models than the data collected afterwards to measure their momentary experiences. In effect the goal is to determine to what extent are we able to predict trait procrastination and momentary procrastination using machine learning models. The time and effort involved in collecting both the data vary significantly (McNeish & Matta, 2018). Different models are used to see the performances of the features. And as the features extracted to infer this premise are collected from the same sample, any deviances in the models' performances could not be associated to the difference in the sample of the population used to conduct this study. This gives us the unique opportunity to understand the innate variance in the features used and the complexities of the models.

One question to be answered right away is why is it important to predict procrastination? Davydov (2014) noted that 64% of employees spend about 2 hours in an 8-hour workday on the internet, surfing through non-work related sites. These costs millions of dollars', worth of productive work, annually for companies. More than the monetary loss, the underlying cause of these procrastinatory behavior could have severe impact such as unmotivated work force, employee satisfaction, loss in creativity and quality of work. Furthermore, engaging in passive procrastinatory behavior, which is not a self-regulatory strategy, is a well-known predecessor of burnout affecting various sects of life (Kumcagiz et al., 2014; Abbasi et al., 2015; Hall et al., 2019).

Psychology as a field is well established in conducting structured experiments in establishing association and relationships between various variables. Over here, the importance lies in predicting future procrastinatory behavior as effectively as possible. Predicting procrastination can help create strategies to increase workforce efficiency (Metin, Peeters and Taris, 2018).When relating this to an academic setting, it can help educators model courses to effectively engage students (Hong et al, 2021). Hence, finding the most efficient way to predict procrastination makes it accessible to everyone and extend it beyond the experimental setup.

This leads to the research questions that this thesis aims to answer.

1.2. Problem Statement & Research Questions :

To what extent does variables from different data collection methods predict trait and momentary procrastination?

The aim of the study is to evaluate the features obtained from different methods of data collection that is used in predicting procrastination. With each method, there is a string of cost involved in obtaining data and processing it. This way it would give us a better understanding of whether the methods used will justify the cost involved and benefits reaped from such tedious procedures. The methods in question are the survey method and the experience sampling method. Answering the following questions will help understand to what degree procrastination can be predicted.

i. *To what extent variables from the survey method predict trait procrastination ?*

ii. *To what extent variables from the ESM predict momentary procrastination ?*

iii. *To what extent variables combined from the survey method and ESM predict momentary procrastination?*

By answering the first two questions we can understand the extent to which machine learning algorithms are able to model different kinds of psychological variables (trait & state variables) when predicting future instances of procrastination. Naturally, the following question would be to see if combing both these trait and state variables produce a better model for prediction.

Different machine learning algorithms are fitted with the variables extracted from the data collected through survey method and ESM. These variables are treated separately since they capture different sorts of information. This way we could try to understand, if the information captured through a certain method prove to be more beneficial in predicting procrastination when fitted with machine learning algorithms. Finally, we combine the variables from both the dataset and analyze if it proves to be more beneficial than treating them as separate entities. Also, the underlying question, what subset of features perform best in predicting procrastination, will be answered.

Answering the above questions will help determine the extent to which one can predict procrastination and provide an understanding of the cost and benefits involved with each method. To answer these questions a variety of regression models, ranging from Linear Regression to Non-linear varieties of regression such as K-Nearest Neighbors, Random Forest, and Extreme Boosting Regressor will be deployed. Different combination of variables from the data collected will be used to model these predictive algorithms.

Main Findings:

Results from the analyses show which predictive algorithms perform better when compared to their respective baseline models. With regards to survey method, Random Forest algorithm with the psychological variable set perform 14% better than the baseline model in predicting trait procrastiantion. ESM variables perform 15% than the baseline model in predicting momentary procrastination using Extreme Boosting Regressor algorithm. They also indicate that psychological features combined from both survey method and ESM predict momentary procrastination better with Extreme Boosting Regressor algorithm. This model's Root Mean Squared Error score on unseen data is 23% better than the baseline model.

To understand these concepts better, the following paragraphs looks deeper into these terminologies.

2. Theoretical Background

2.1. Procrastination

Procrastination is essentially seen as a modern-day malady with industrious scheduling playing an important role in compartmentalizing our day-to-day activities (Milgram, 1992). To understand the importance and ill-effects of procrastination, we need to first define it. Ferrari, Johnson & Mccown (1995) define procrastination as a dilatory behavior. Dilatory is associated with a negative connotation, such as, lazy, tardy, sluggish, etc. But does this association have to be extended to procrastination as well? If procrastination is such a negative behavior, why are most people guilty of doing it?

People often procrastinate when trying to delay the start or completion of a particular activity (Steel, 2007). It is commonly associated with the instant gratification principle where one fixes their attention to activities that gives them instantaneous pleasure rather than performing a task that needs to be completed. Let's take an example where a student engages in watching the recent YouTube videos of John Oliver instead of completing assignments. This triggers the pleasure response cycle and reinforces such a behavior and before they know it, they are in a spiral of watching random videos and wasting essential time. Zhang & Feng (2020) explained in their paper on the temporal decision-making model that people would procrastinate as long as the aversiveness of a task outweighs the benefits of completing it.

Before going deeper into the impacts of procrastination, it is important to differentiate between Trait procrastination and Momentary procrastination. Reinecke et al., (2018) defines trait procrastination as the unchanging inter-individual differences in the manifestation of dilatory behavior across different life domains. This is how one characterizes themselves as a procrastinator, meaning an average estimation of one's procrastination level. Research clearly demarcates such differences are in part stemming from one's genetic composition, such as their impulsivity (Gustavson et al., 2014) and in their personality traits (Schouwenburg & Lay 1995; Steel, 2007). In contrast to trait procrastination, momentary procrastination is one's reaction to a specific task (Schouwenburg, 2004). This is essentially the manifestation of dilatory behavior in response to task. This could be in response to a task which is perceived as unpleasant, difficult, aversive, ambiguous, stressful, etc.

Procrastinatory behavior comes with its cost and benefits. In academic settings, procrastinators are correlated to show poor academic results (Tice and Baumeister, 1997; Kármen et al., 2015; Rajapakshe, 2021). This is stipulated due to the procrastinators inability to manage time effectively, thereby inducing time pressure and stress (Choi and Moran, 2009). This in turn translates to lower quality of work and poor grades. Looking into the cost of procrastination at workplace, it is estimated that an average employee loses up to two hours each day just on surfing the internet for personal use, socializing with colleagues, etc. causing loss in productivity (Malachowski & Simonini, 2006; D'Abate & Eddy, 2007). Ferrari (2001) notes that when "working under pressure", most procrastinators work slowly and make more errors than non-procrastinators.

On the other hand, Steel (2007) also refers to procrastination as a functional delay which helps in information gathering and evidence building when performing important tasks (Ramsay, 2006). When procrastination is implemented as a micro-break strategy, performance seems to be better (Tice and Baumeister, 1997; Chauhan et al., 2020). Thus, procrastination as a behavioral trait is notoriously expensive and needs to be appropriately identified to avoid the ill-effects caused by it. Whereas procrastination when used as a self-regulatory strategy seems to alleviate stress and manage time effectively.

2.2. Procrastination and Survey method

Survey research is one of the most popular methods of data collection in social sciences. It is the use of standardized questionnaires or interviews to collect data about people's behavior, thoughts, or preferences in a systematic way (Bhattacherjee, 2012). While its main usage is in quantitative research, it can also be employed in descriptive, exploratory, or explanatory research. It is an excellent way to gather information about unobservable data variables, such as income, self-esteem, attitude towards certain groups, etc.

A major part of research about procrastination is done through the survey method (Malachowski & Simonini, 2006; D'Abate & Eddy, 2007; Chuan et al., 2020; Rajapakshe, 2021). Schouwenburg & Lay (1995) compared in their study the relationship between trait procrastination and personality factors in Dutch and American students. They stated that there is a negative correlation between Procrastination and Conscientiousness and a positive linear relationship between Procrastination and Neuroticism. They suggested that when one feels their environment to be distressing and unsafe, they tend to escape these feelings by focusing on something less threatening. This escape mechanism leads to procrastination. These findings were in line with study done McCown et al. (1987), but McCown also observed evidence of non-linear relationship between trait procrastination and neuroticism.

When looking into the relationship with psychological factors, procrastination is seen to have a positive correlation with perceived stress (Tice & Baumeister, 1997). Whether stress is the outcome of procrastination or higher stress levels brings about procrastination is still debated. When it comes to fatigue, Gropel & Steel (2008) noted in their study a positive correlation between procrastination and fatigue. This means that when people engaged in dilatory behavior, they felt more mentally exhausted. This goes hand in hand with escape mechanism. When one constantly tries to avoid a certain task actively, they are using up mental resources to engage themselves negatively. This shows that high procrastinators are susceptible to decreased mental health.

All the studies mentioned above have one thing in common. It is the way the data was collected. These studies have their participants fill in standardized questionnaires for quantitative analysis. This one-shot method makes it feasible to attain data from a large sample of population remotely without having to observe them directly. Also, several variables could be measured at the same time. With the advancement in technology, online surveys make it easier to reach a wider audience more easily. Moreover, they are also cost effective (Das, 2012).

As it is with all methods, the survey method has its flaws. Recall bias is one of its major concerns (Podskaoff et al., 2003). Recall bias is an information bias that occurs when there is a discrepancy between what the subject reports and what the actual truth is. A good example would be questions asking them to report their levels of stress or anxiety a month ago. There is a good chance that they do not recall that information accurately. Either they inflate their responses depending on how they are feeling at that moment or deflate their responses wanting to appear socially desirable (Bhattacherjee, 2012). Another problem is the response rate. An ideal length of a survey would be 15 to maximum 20 mins (Bhattacherjee, 2012). If it is longer than that participants start to lose their interest and this increases the chances of recall bias.

2.3. Procrastination and ESM

Experience Sampling Method is often preferred by researchers with an interest in studying human behavior (McNeish & Matta, 2018). Participants of this study type are requested to self - report their behavior (whatever the study requires) multiple times a day. They note their behavior by answering a short, usually identical questionnaire which are prompted by notifications sent to their mobile phones. These reports are collected over multiple days from several participants. This gives researchers an opportunity to study behavior in the participants' natural environment, as opposed to in a laboratory. By this way, an accurate representation of the subject's behavior could be noted.

In the context of procrastination, ESM is used in a variety of different frameworks. Aalbers et al., (2021) studies the relationship between procrastination and passively logged smartphone use data using ESM. In their study they noted a positive correlation between momentary (state) procrastination and smartphone notifications, total smartphone usage and use of specific smartphone use categories. That is when a person received a notification, they had higher tendencies to engage in this dilatory behavior than when they did not. And the same could be said about the smart phone use as well.

When one logged in higher smartphone use in a particular window between them filling their questionnaires, they also mentioned to have noted higher scores in procrastination. This is empirically verified by collecting data using ESM.

Wessel, Bradley, & Hood (2019) uses ESM to measure behavioral delays in situ to understand the differences between active and passive procrastinators. This longitudinal study shows a strong positive correlation between scores from passive procrastination scale and averaged behavioral delays. In another study, after measuring delays associated with procrastination, Wessel et al., (2020) administered low intensity, high frequency intervention successfully to alleviate procrastination through ESM. That is, they administered an intervention program to high procrastinators to help them manage it. This method of data collection has proven to be useful in its flexibility to encompass a variety of facets that otherwise would prove to be difficult to implement.

ESM on the other hand is quite expensive to implement. Expensive in the sense of participant burden, programming the mobile application, study cost, software compatibility, etc. (McNeish & Matta, 2018). They also stress upon the fact that, skills required to implement such a study is still high and that the quality of data obtained is still questionable. As the data is collected longitudinally in a remote setting, the mutual alliance between the researcher and the subject is lost. This in turn could lead to poor quality of data.

2.4.Procrastination and Machine Learning

Yarkoni & Westfall (2017) in their paper describe how scientific psychology has merged explaining behavior and predicting behavior as a single process. Philosophically they may be compatible in the sense that when building the best model to explain a certain phenomenon, it is possible to use the same model to predict future instances. This notion is refuted from a statistical point of view. It is often not true that a model that closely estimates all the data points from "data-generating process" (in simple words the training data) will also be successful at predicting unseen data (Shmueli, 2010). Creating models to predict unseen data to its best is the gold standard of machine learning.

"What is machine learning?" is a natural question to follow. Daume (2012) defines machine learning in simple terms as, "predicting future behavior based on the past instances". Predicting if 'person A' likes an unseen new movie based on the ratings they had given for movies in the past is a classic example of machine learning. This is what recommender system on Netflix is built with. The factors contributing to this prediction vary depending upon the problem at hand.

When it comes to procrastination, machine learning techniques has been used in some interesting studies. Xu (2020) in their study analyzed twitter data with hashtag “#procrastination” to find it to be in line with self-enhancement strategy, that is, when people tend to represent themselves on social media it is usually in a positive way. A trained machine learning classifier is used to automatically classify the tweets into several categories. Their tweets were found mostly to be in a positive tone despite their ironic representation.

In another study, Linear Support Vectors Machine and Neural Networks are used to classify students’ performance through their procrastination behaviors (Hooshyar et al., 2019). They used educational data mining techniques to include their past academic data along with non-academic factors to predict students’ academic performance. Several more studies have seen the use of classification models to predict submission of assignment (Dragulescu et al., 2015), predicting student’s procrastination through their grades (Akram et al., 2019) and classifiying students at risk of high procrastination through assignment submission pattern (Olive et al., 2019).

The studies detailed above have shown good results in constructing classification models. They have utilized the advantages of using machine learning models to form relationships between variables which are hard to otherwise infer and analyze. However, how far these variables are generalizable is still a question. These studies have considered attributes that are subjective to their situation which makes it difficult to implement and interpret. These studies show a pattern where procrastination has been not measured with a standardized scale but rather inferred through other attributes such as lack of engagement, delay in assignment submission, etc. There seems to be lack of studies including machine learning models in predicting psychological factors.

To bridge this gap, this study combines the psychological factors measured through standardized questionnaires to see how far these variables could be used in predicting procrastination variable. This study also uses some non-behavioral features along with these psychological features to see their interplay using different regression algorithms. It also uses various variables that are theoretically sound in predicting procrastination and will be put in action to see how well they perform on unseen data.

3. Experimental Setup

3.1. Dataset description

3.1.1. Data.csv:

For this study, the dataset (Aalbers et. al., 2020) in Comma Separated Value (CSV) file is used. It contains information about the participants from Tilburg University who volunteered in a study regarding smartphone usage. This has records of participants responses from the behavioral scales. These scales can be divided into three categories, the initial onboarding survey, the monthly survey, and the daily experience sampling survey. The responses to these questions were concatenated together into one dataset.

The raw dataset came with 69,880 rows and 160 columns. The dataset consists of 236 participants with android phones. The demographics include Tilburg University students of mean age 20.49 years with 129 female and 107 male participants. Along with age, sex (biological) and gender (which sex they identify with), each question's response were noted separately for each scale. For example, Big Five personality scale consists of 30 questions. The response to each question were noted separately. The Appendix A shows the list of the column names included in the dataset.

3.2. Data preprocessing:

Data preprocessing here primarily included getting total score of each variable and separating the onboarding survey data from the experience sampling surveys. Each column was investigated separately to check if there were any erroneous entries present. This was done by using boxplots to see if the outliers were valid or not. For example, from Fig. 3.2.1, User #23926 age was found to be -0.41. This for sure is not possible, hence this user was removed from the analysis i.e., 302 entries were dropped. Also, 18,417 entries were dropped due to missing information from the ESM surveys. Participants were encouraged to fill in the experience sampling questionnaire 5 times a day. If they had failed to complete a particular survey, that slot is filled in as missing information. This study is treating each ESM entry as an individual entity. By doing so we model the relationship between the dependent and independent variable at an individual level excluding any within person or between person effects. Hence, all the missing entries were removed. This ended with a total of 51,161 rows of data from the original 69,880 rows of data.

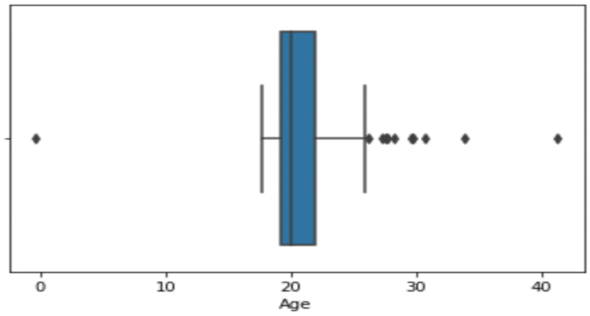


Fig 3.2.1 Boxplot of Variable Age

Once the dataset is cleaned, it is then separated into 3 different datasets. The first one is 'survey_method' dataset, which contains all the features of collected during the initial onboarding survey. This dataset contains 235 rows and 17 columns. The next is 'ESM_method' dataset which consists of 51160 rows and 16 columns. This dataset has the ESM variables along with duration it took to fill in the daily survey, type of the survey, day of the week it was filled in along with the demographics of the participants. The third dataset consists of both the survey and ESM variables cleaned and combined. This resulted in 51160 rows and 27 columns. I chose to make three new csv files from the original dataset to facilitate easier analysis.

How the outcome variable and other features were extracted from the initial dataset is explained in the following sections.

3.2.1. Outcome variable:

In the survey method, procrastination was measured using the General Procrastination Scale - Screening (Klein et al., 2017). The GPS-S consists of 5 items on a 5-point Likert scale (1 - not at all, 3 - moderately, 5 - extremely). The responses of this scale were noted in columns 'PRO1', 'PRO2', 'PRO3', 'PRO4' and 'PRO5'. The items were preceded by the statement "To what extent these statements apply to you in general". Participants were asked to note down in general if they engaged in activities that leaned towards procrastination. For example, "In preparation for some deadlines, I often waste time by doing other things". The average of the responses recorded from these 5 items is used as the procrastination score of the participants.

As for the ESM, General Procrastination Scale - Experience Sampling Method is used. GPS-ESM is based on the GPS-S from Klein et al., (2017). This scale is of 3 items on a 7-point Likert scale (1- not at all, 4 - moderately, 7 - very much). The responses to this scale were noted in columns 'P1', 'P2' and 'P3'. These items were preceded by the statement "Please report to what extent the following statements applied to you since the last survey". The items were presented in a fixed order: "I delayed before starting on work I have to do", "I wasted time by doing other things than what I had intended to do", and "I thought: 'I'll do it later.'". For 30 days, Ethica sent the participants notifications 5 times a day to day at pseudo - random time slots between 8.30am and 10.30pm to complete these brief surveys. The average of the responses to these 3 questions is used as procrastination score for the ESM study.

3.2.2. Features:

Features used in this study are motivated through literature. The features used in the survey method include,

1. Big-five personality: To measure the participants personality traits, BFI-2- short form (Soto & John, 2017) was used. This scale contains 30 items. The five traits that is measured are Extraversion ('BFT1', 'BFT6', 'BFT11', 'BFT16', 'BFT21', 'BFT26), Agreeableness ('BFT2', 'BFT7', 'BFT12', 'BFT17', 'BFT22', 'BFT27'), Conscientiousness ('BFT3', 'BFT8', 'BFT13', 'BFT18', 'BFT23', 'BFT28'), Neuroticism('BFT4', 'BFT9', 'BFT14', 'BFT19', 'BFT24', 'BFT29'), Openness ('BFT5', 'BFT10', 'BFT15', 'BFT20', 'BFT25', 'BFT30'). These traits were treated as separate continuous features and used in prediction. Also, the traits were clubbed together and used as one score, and this was a separate feature used in prediction. One model contained all the personality traits separately and another model contained one score for personality.

2. Perceived stress: Perceived Stress was measured using the PSS - 10 (Roberti, Harrington & Storch, 2017). This 10-item scale captured two specific dimensions of perceived stress, which are the perceived helplessness ('PSS1', 'PSS2', 'PSS3', 'PSS6', 'PSS9', 'PSS10') and perceived self-efficacy ('PSS4', 'PSS5', 'PSS7', 'PSS8'). Similar to personality scores, one model used the dimensions of perceived stress and other model used stress score averaged across both the dimensions.

3. Fatigue: MFI from Smets et al., (1995) was used to measure fatigue levels in participants. This scale consists of 20 items. The average over these 20 items is used as the final fatigue score. ('FAT1' till 'FAT20').

4. Connectedness: Social connectedness is measured using an 8 - item scale. SCS scale assesses one's level of need for belongingness. This scale is developed by Lee & Robbins (1995). An average over items 'CON1' to 'CON8' gives the score of social connectedness.

5. Social Desirability: The level of social desirability is assessed using SDRS - 5 (Hays, Hayashi & Stewart, 1989). This is a five-item scale that measures the level of pressure one feels to behave in a socially approved manner or feel socially approved feelings (Hays et al., 1989). Items 'SOD1' to 'SOD5' are averaged to get the final social desirability score.

6. Impulsivity: Impulsivity is measured using Brief BIS - 11 scale (Morean et al., 2014). This is an 8 - item scale. The responses were averaged to get the final impulsivity score. ('IMP1' to 'IMP8').

The features used in the ESM study include,

1. Stress: To note the momentary experience of the participants, stress was measured using 3 item scale with a 7-point Likert scale(1 - "not at all", 4 - "moderately", 7 - "very much"). The questions included were "I feel rushed", "I feel relaxed", "I feel stressed (tense, restless, nervous, or anxious). The average of these responses was used as stress scores for the ESM.

2. Fatigue: A procedure similar to "Stress" was implemented for fatigue. The items included here were, "I have enough energy", "I feel desire to do things" and "I can concentrate well".

3. Happiness: One question was used to determine the participants level of happiness at that moment on a 7-point Likert scale. This question was, "How happy do you feel right now?"

Apart from these behavioral features, other factors were included to see if they improve the models predicting momentary procrastination. These were,

1. Type: This refers to the type of the ESM survey participants filled each day. There are three types of survey. First is 'M' which refers to the first survey they fill each morning. Next is 'D' which refers to the other four surveys they fill throughout the day. If they missed the 'D' surveys, they had 'E' survey, which is the emergency survey they filled for the day.

2. Duration: Duration is how long participants took to complete the survey. This is measured in seconds.

3. Day of the week: From the date when the surveys were issued, the day of the week was noted using the 'pandas' function "`df.dt.dayofweek`" where df stands for the data frame.

3.3. Exploratory Data Analysis:

Exploratory Data Analysis (EDA) is the first step in the data analysis process. It involves the examination of patterns, trends, outliers, and unexpected results in the dataset. One important point to note is that the variables in the survey method range from 1 to 5 whereas variables from the ESM range from 1 to 7 on the Likert Scale. The preliminary EDA shows that most of the variables from the survey method are normally distributed. This cannot be said about the outcome variables of ESM and Survey Method procrastination. This does not affect the use of regression algorithms to model these variables. Fig 3.3.1 shows the distribution of outcome variable from the survey method.

It is not possible to compare these two variables as they are capturing different information (one captures trait overall and the other captures one's current state). If they need to be compared, it is essential to average the responses across each participant in the ESM data. This leads to loss of important information. Therefore, we treat them as two separate entities and observe their behavior across various machine learning models. This gives us insight into how feasible it is to predict each category of procrastination with least amount of costs involved.

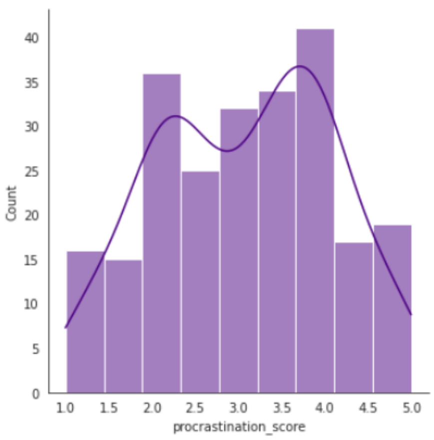


Fig. 3.3.1. Density plot SM procrastination

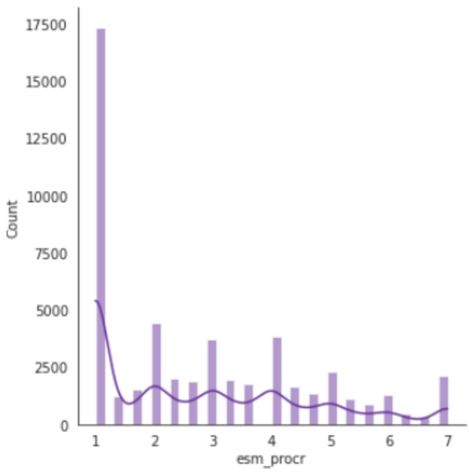


Fig.3.3.2. Density plot ESM procrastination

The density plots of the behavioral features have been added to the appendix. It is interest to observe that the ESM features show a unique pattern. This could also indicate an information imbalance in the dataset (Yang et al., 2021). There is a lot of information available about a certain datapoint rather than the other. Here if you look at Fig. 3.3.2. it shows a lot of information available about when the procrastination score is 1, compared to the other scores, which makes it difficult to train a machine learning model effectively.

The tables 3.3.1 and 3.3.2 shows the means and standard deviations of the variables. The standard deviation shows the amount of variation or dispersion in a dataset. A small standard deviation (SD) indicates that the values are closer to the mean and large SD indicates that the values are farther away from the mean. As you can see, the mean of the ESM procrastination is 2.79 when the range of this value is from 1 to 7. When a machine learning model just predicts the mean continuously, it will end up with a higher error score than when compared to SM procrastination.

Table 3.3.1 Mean & SD of Survey Method Features

Variable	Mean	SD
Age	20.49	2.80
Extraversion	3.14	0.31
Conscientiousness	3.28	0.40
Neuroticism	3.24	0.40
Openness	3.16	0.36
Helplessness	2.86	0.80
Self - Efficacy	3.37	0.70
Fatigue	2.89	0.20
Connectedness	2.55	1.12
Impulsivity	2.82	0.31
Social Desirability	2.85	0.50
Procrastination	3.07	1.01

Table 3.3.2 Mean & SD of ESM Features

Variable	Mean	SD
Duration	719.65 (in secs)	825.29 (in secs)
ESM Fatigue	3.95	1.29
ESM Stress	3.33	0.76
ESM Happiness	4.63	1.37
ESM Procrastination	2.79	1.80

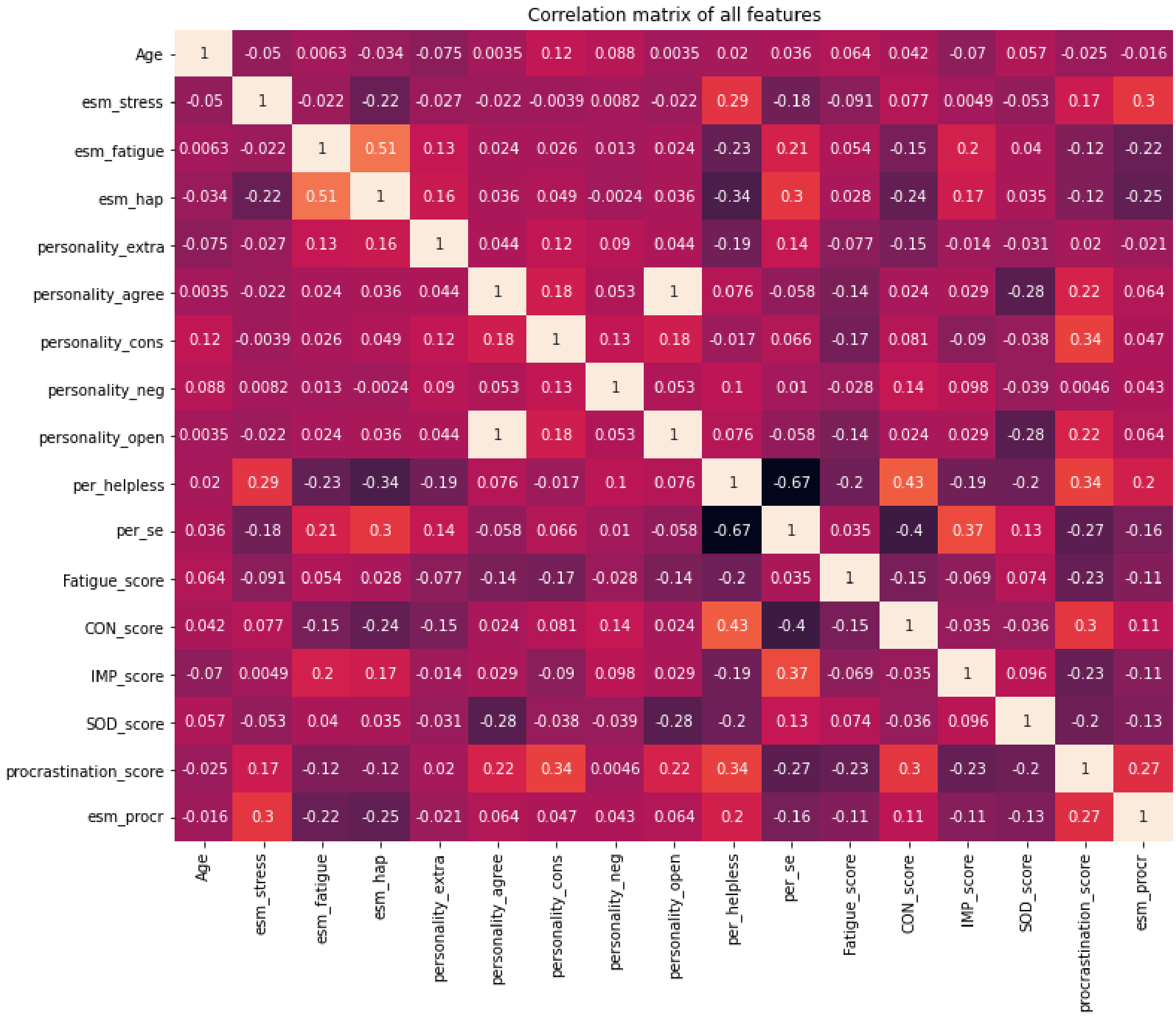
Table 3.3.3 Value counts of ESM Features

Variable	**Counts**
Monday	7298
Tuesday	7459
Wednesday	7517
Thursday	7601
Friday	7438
Saturday	6949
Sunday	6898
Type D	38692
Type M	7568
Type E	4900

Other than the behavioral features, information about the day of the week when the ESM questionnaire was filled is also noted. Along with it the Type of ESM survey is also used. Their counts are mentioned in table 3.3.3.

In this study, the variable Gender will be used to see if it contributes to the regression models. Gender is used instead of Sex as it describes how one identifies oneself rather than their biological disposition. User 23833 is born a Female but identifies oneself as Male. User 25273 is born a Male and identifies oneself as Female. User 24346 and 25180 are both born as Male but rather does identifies as non-binary.

During the Exploratory Data Analysis, the correlations between the variables were checked. The following heat map is made using the ‘`seaborn`’ module, show the relationship between the variables.



The above map shows poor to weak correlation between the independent variables and procrastination variables from both the Survey Method and ESM. Correlation coefficients with +/- 1 is considered ‘Perfect relationship’, +/- 0.7 to 0.9 is considered ‘Strong relationship’, +/- 0.4 to 0.6 is considered ‘Moderate relationship’, +/- 0.1 to 0.3 is considered ‘Weak relationship’ (Schober, Boer, & Schwarte, 2018). Pearsons correlation measures the linear relationship between variables. From the map above it is noted that in the variables from the survey method have a weak linear relationship with procrastination and a similar pattern is noted between ESM procrastination and other ESM variables as well.

Conscientiousness, perceived helplessness (from perceived stress scores) and connectedness has a weak linear relationship (r = 0.34) with trait procrastination. This is in line with previous studies. There seems to be weak relationship even between other personality traits such as agreeableness, openness with trait procrastination. What is interesting is also a weak linear relationship between trait procrastination and momentary procrastination. This is understandable as participants who scored high on trait procrastination could have been inclined towards scoring higher momentary procrastination scores as well.

There are intercorrelations between multiple independent variables. There is a moderate positive relationship between ESM Fatigue and ESM Happiness. Previous studies (Indhira & Shani, 2016; Kwon, 2020) suggests more a negative relation between these variables. There is also a high correlation between perceived helplessness and perceived self-efficacy. This is understandable as they are different facets of the same scale which is the perceive stress scale (PSS). These moderate to weak correlations between the independent variables might affect the interpretability of the regression algorithms but it does affect the predictions nor the precision of the predictions (Kutner et al., 2004). Hence, this study uses linear and nonlinear methods to analyze which combination of variables explains the variability best in procrastination (both ESM and SM).

3.4. Algorithms:

In this section, we talk about the different regression algorithms used in this study.

3.4.1. Regression models:

Wopert et al., (1997) states in their "No Free Lunch" theorem that there is no one optimal machine learning algorithm for any problem. Every dataset has its own unique challenges. Therefore, the best suited model is highly dependent on the problem at hand. In this study, various regression algorithms have been used to find the best way to predict procrastination. Also, to take into consideration the uniqueness of each dataset, and the challenges that comes with it, a series of linear and non-linear models are used.

We are considering here supervised learning models since we have both the independent and dependent variable available for training. As both the procrastination scores are continuous variable, regression algorithms are used to model the dataset to find the best fit to predict procrastination.

<u>Baseline model:</u>

To assess how well the algorithms are performing, a baseline model is required. A baseline model is a simple, easy to implement algorithm that serves as a starting point. Since there is not much research on using machine learning models in predicting psychological variables, dummy regressor from '`sklearn.dummy.DummyRegressor`' is used as the baseline model. This algorithm makes predictions using simple, straightforward rules. The strategy used here for the algorithm is "mean", i.e., it always predicts mean of the training targets.

Linear Regression :

Linear Regression is one of the classic learning algorithms used in prediction problems. Regression is a method where the target variable is modelled based on the relationship between the variables (independent and dependent). Linear Regression assumes a linear relationship between these variables. Even though, both our datasets showed only weak correlations, Linear Regression is one of the most preferred algorithms to predict behavioral features. Also, various studies have shown linear relationship between the number of independent variables that are available in this study with procrastination (D'Abate & Eddy, 2007; Sirois 2014; Chuan et al., 2020; Rajapakshe, 2021). Hence, we use linear regression as to see if it pertains to any of the dataset used in this study. It is implemented using '`LinearRegression`' algorithm from '`sklearn.linear_model`' module.

K-Nearest Neighbor :

K - Nearest Neighbor (k-NN) is a non-parametric learning algorithm which is used for both regression and classification problems (Altman, 1992). K-NN regression in an insightful approach, it approximates the connection between the predictor variables and the continuous target variable. This is done by averaging the outcomes that are in the same neighborhood. Hence, setting up the 'k' hyperparameter in k-NN is crucial as it helps to select that size that minimizes the mean squared error.

K-NN would serve as a good algorithm to consider as it helps to contain the problem of curse of dimensionality (Kpotufe, 2011). As data relating to self-reports are highly irregular, it might help to combat the problem of high dimensionality by being adaptive to low intrinsic dimensional manifold. This model is implemented with '`sklearn.neighbors`' module's '`KNeighborsRegressor`'.

Random Forest Regression :

Random Forest is an ensemble method that combines multiple models to get a stronger model (Varghese, 2018). With respect to Random Forest, multiple decision trees are combined. A decision tree is hierarchical and has a tree-like model of decisions and their possible outcomes. It chooses as the first input predictor the one that explains the most variance in the target variable. With a Random Forest the decision trees are fed random input variables, this is a measure to reduce the variance (James, et al., 2017). This results in a more robust and accurate model which handles overfitting better than its fundamental model, which is a single Decision Tree. This algorithm's nonlinear nature would help to fit the datasets better. Random Forest is implemented using '`RandomForestRegressor`' from '`sklearn.ensemble`'.

XGBoost :

Extreme Gradient Boosting is a go to model for machine learning competitions (Brownlee, 2021). Gradient Boosting a class of ensemble models that is used in all sorts of predictive modelling, i.e., regression and classification. XGBoost is known for its quick implementation and accurate model execution (Brownlee, 2021). This model is also very good at incorporating datasets with both continuous and categorical variables. Like Random Forest, this is also a tree-based algorithm that uses an ensemble of decision tress. Unlike Random Forest, it builds its tree one at a time thereby repetitively leverages patterns in its residuals to build a stronger model (Glen, 2018). Hence, this algorithm would fit even better with the datasets. This algorithm is implemented using '`xgboost`' module with '`XGBRegressor`'.

All the above algorithms were implemented after scaling the features using '`StandardScalar`' function from '`sklearn.preprocessing`'. A pipeline was created using the function '`Pipeline`' from the module '`sklearn.pipeline`'. In machine learning, pipeline is a way to automate the workflow to produce a machine learning model. With this function the features are first scaled and then the regression model is built sequentially. This is essential for liner regression and KNN.

3.6. Evaluation:

To train the models and find the best fitting hyperparameters , 5-fold cross validation is used. Cross validation is primarily used to see how well the model performs on unseen data . And to use 5-fold cross validation, it helps to keep in check that the model does not overfit on the training data (Refaeilzadeh, Tang, & Liu 2009).

Root Mean Squared Error is the evaluation metric used to see how well the model is performing on the training, validation, and test data. RMSE measures the standard deviation of the residuals. Residuals show how far the prediction is from the true target value. They show a sample of the true error of the population. Therefore, the RMSE scores show us magnitude of the expected prediction error.

3.5. Hyperparameter Tuning and selection:

Learning algorithms have several hyperparameters that could be modified. Hyperparameters are values that is used to control the learning process in an algorithm. Modifying these parameters could yield in improved performance of the algorithm. To find the best combination of the parameters for each algorithm '`GridsearchCV`' from '`sklearn.model_selection`' is used. This function fine tunes the parameters systematically with pre-set values. The data is split into 75% for the training and validation set and 25% for the test set. This is done so that the tuned hyperparameters are later evaluated on unseen data to check for model's genralizability. The GridSearchCV splits the training data further with the use of cross validation into training and validation sets. This step is repeated 5 times where before each split the data set is shuffled. This gives us more optimized tuning and helps to prevent overfitting. The following table shows the hyperparameters that were tuned for KNN, Random Forest and XGBoost algorithms.

Table 3.5.1. Hyperparameters tuned in each algorithm

MODEL	**HYPERPARAMETER**
KNN	K: number of neighbours
Random Forest	max_depth: maximum depth a tree can have n_estimators: number of trees
XGBoost	max_depth: maximum depth a tree can have learning_rate: how quickly the error is corrected from one tree to next

3.7. Software

In this study, Python is used to build models and perform analyses. This is done using Jupyter Notebook hosted on Google Colaboratory servers . Several packages within Python are used to pre-process and model data. For pre-processing, 'Numpy' and 'Pandas' packages have been used. To construct models for predictions, 'Sklearn' packages have been used. Finally, 'Matplotlib' and 'Seaborn' packages have been used for data visualization.

Packages	**Version**	**Source**
NumPy	1.20.0	Harris, C.R., et al. (2020)
Pandas	1.2.4	McKinney, Wes, & others (2010)
Scikit-learn/ Sklearn	0.24.2	Pedregosa, et al., (2011)
Matplotlib	3.4.1	Hunter (2007)
Seaborn	0.11.1	Waskom, et al. (2018)
XGBoost	1.4.1	Chen, T., & Guestrin, C. (2016)
Scipy	1.7.0	Virtanen et al., (2020)

4. Setup

For this regression problem, different sets of feature values are used. Each set is a unique combination of the features extracted from the dataset. For the Survey method, four sets of features are used. The table 4.1 outlines the sets of models used in this study. Set 1 includes all the psychological trait variables and demographic variable available in this study. As for Set 2, the dimensionality of personality and stress scores are expanded and uses each facet as an independent variable along with age and gender. Set 3 is Set 1 without age and gender. As the distribution of age in the participant pool is not that varied, Set 3 would check if there would be a depreciation in the results due to removing these two variables. Set 4 uses the variables from Set 2 minus age and gender variables.

Table 4.1 Sets of features examined from Survey Method

Set number	Features
Set 1	Age, Gender, Personality, Stress, Fatigue, Social Connectedness, Impulsivity, Social Desirability
Set 2	Age, Gender, Extraversion, Agreeableness, Conscientiousness, Openness, Neuroticism, Helplessness, Self-Efficacy, Fatigue, Social Connectedness, Impulsivity, Social Desirability
Set 3	Set 1 without Age and Gender
Set 4	Set 2 without Age and Gender

For the ESM a similar pattern is followed. Table 4.2 shows the sets of features used while predicting ESM procrastination. Set 1 includes all the features from ESM including the duration took to fill the daily surveys, the day of the week and the type of the survey. Model 2 excludes the type of the survey and the day of the week. Model 3 excludes the Age & gender. Model 4 consists only of the items from the daily survey. By systematically removing features, we see if there is a dip or raise in RMSE scores with each feature set.

Table 4.2. Sets of features examined in ESM

Set number	Features
Set 1	Age, Gender, Duration, Stress, Fatigue, Happiness, Day of the week, Type
Set 2	Age, Gender, Duration, Stress, Fatigue, Happiness
Set 3	Duration, Stress, Fatigue, Happiness
Set 4	Stress, Fatigue, Happiness

In the mixed model setup, four sets of features are compared to see which combination provides the best RMSE scores. Table 4.3 gives the outline of it. To run models with mixed features (features from both survey method and ESM), the dependent variable is the procrastination score from ESM. As the ESM procrastination variable chronologically occurs last, the study will run models to see if one's general behavior traits along with their momentary levels of stress, fatigue, happiness affects their momentary levels of procrastination. Set 1 displays the full range of features from both the methods with survey methods dimensionality of personality and stress scores expanded. Set 2 has the condensed version of these two scores. Set 3 and 4 are versions of set 1 and 2 without non behavioral features in it.

Table 4.3. Sets of features examined in Mixed model

Set number	Features
Set 1	Age, Gender, Duration, ESM_Stress, ESM_Fatigue, ESM_Happiness, Day of the week, Type, Extraversion, Agreeableness, Conscientiousness, Openness, Neuroticism, Helplessness, Self-Efficacy, Fatigue, Social Connectedness, Impulsivity, Social Desirability
Set 2	Age, Gender, Duration, ESM_Stress, ESM_Fatigue, ESM_Happiness, Day of the week, Personality, Stress, Fatigue, Social Connectedness, Impulsivity, Social Desirability
Set 3	Set 1 without Age, Gender, Duration, Day of the week, Type
Set 4	Set 2 without Age, Gender, Duration, Day of the week, Type

5. Results

This section will provide the results from the various analyses. First, the results from the survey method are presented followed by the ESM model results and finally the mixed model results.

5.1. Survey method :

In the survey method, the responses from the onboarding interview were used to see if trait procrastination could be predicted. The **baseline model RMSE is 1.017**. This was calculated using the Dummy Regressor method. The mean of the procrastination scores were predicted continuously and how far these responses were from the true target value was calculated as the RMSE.

The tuned hyperparameters used for KNN, Random Forest and XGB Regressor are shown in Table 5.1.1 and the Test RMSE of each sets' performance for the models used are shown in Table 5.1.2.

Table 5.1.1. Optimal Hyperparameters

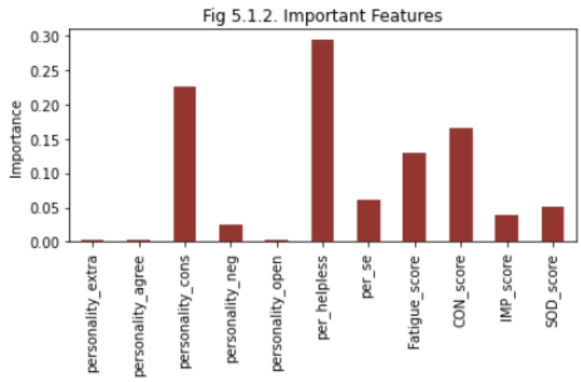
Set No.	KNN	Random Forest	XGB
Set 1	12	max_depth = 3, n_estimators = 300	max_depth = 2, learning_rate = 0.05
Set 2	15	max_depth = 3, n_estimators = 300	max_depth = 1, learning_rate = 0.1
Set 3	16	max_depth = 4, n_estimators = 300	max_depth = 1, learning_rate = 0.05
Set 4	20	max_depth = 2, n_estimators = 50	max_depth = 1, learning_rate = 0.05

Fig.5.1.1. Training RMSE of Survey Method

This classic way of data collection measured the trait variables. Modelling these variables resulted in the following prediction results. Out of 16 sets of combinations of models 13 sets performed better than the baseline model. The three that did not are Set 2 in Linear Regression and Set 1and Set 3 in XGB Regressor. Set 4 with Random Forest performed the best. Most of the sets with tree-based algorithms shows a minor overfitting on the training data. Set 1 to 3 on Random Forest and Set1 to 4 on XGB Regressor show overfitting. Set 4 with Random Forest does not overfit on Training data. The RMSE score of the training data is 0.805 and Test RMSE is 0.87. This combination performs 14.5% better than the baseline model.

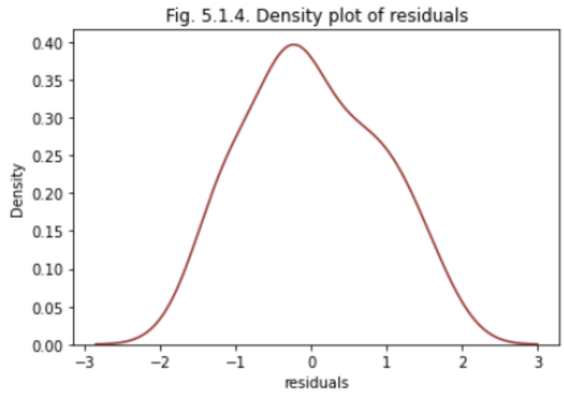
Table 5.1.2. Test RMSE of Survey Method

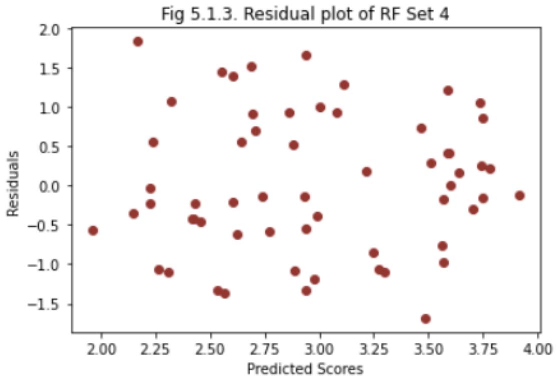
Set Number	Linear Regression	KNN	Random Forest	XGB
Set 1	0.91	0.91	0.87	1.07
Set 2	1.07	0.91	0.94	0.97
Set 3	0.95	0.93	0.98	1.08
Set 4	0.99	0.92	**0.87**	0.91



It is accurate to say that this model does not capture all the variance in the dependent variable. The variables that contribute the most in predicting procrastination are shown in Fig.5.1.2. ‘Perceived helplessness’ contributes the most followed by the personality trait ‘Conscientiousness’. ‘Social connectedness’ and ‘Fatigue’ scores are the next biggest contributors. There are also small contributions from ‘perceived self-efficacy’, ‘social desirability’, ‘impulsivity’ and personality trait ‘neuroticism’. Other personality traits such as ‘extraversion’, ‘agreeableness’ and ‘openness’ do not contribute to this regression model.

The residuals of the out of sample data are normally distributed. Fig 5.1.4. shows the normal distribution of these residuals and Fig.5.1.3. shows the residuals plotted against the predicted





values of Set 4. The scatter pattern shows that the residuals do not have any patterns. Also, one thing to note is that these residuals are placed further away from 0 than near it. It could possibly mean that the model is missing key information when trying to predict procrastination or needs more instances.

5.2. ESM :

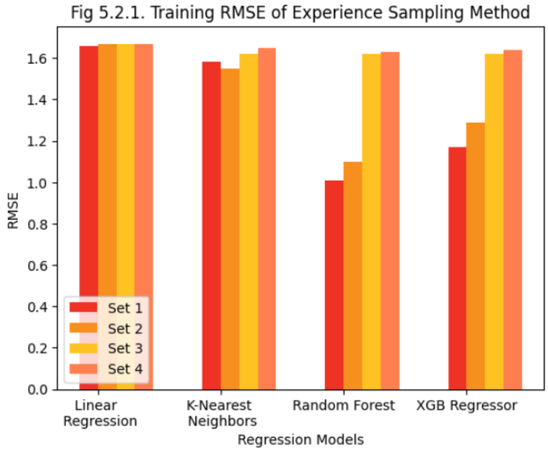
In this section, the results from the ESM models are noted. The baseline is calculated like the baseline in Survey Method. **The baseline model RMSE is 1.805**. The baseline RMSE for ESM method is higher since the procrastination variable has a range from 1 to 7 on the Likert scale with a mean of 2.79.

The optimal hyperparameters used for KNN, Random Forest and XGB Regressor are shown in Table 5.2.1.

Table 5.2.1. Optimal Hyperparameters

Set No.	KNN	Random Forest	XGB
Set 1	28	max_depth = 15, n_estimators = 300	max_depth = 9, learning_rate = 0.09
Set 2	24	max_depth = 14, n_estimators = 300	max_depth = 8, learning_rate = 0.09
Set 3	32	max_depth = 7, n_estimators = 200	max_depth = 4, learning_rate = 0.1
Set 4	236	max_depth = 7, n_estimators = 300	max_depth = 4, learning_rate = 0.1

Fig 5.2.1. Training RMSE of Experience Sampling Method



The trained models' performance on unseen data is noted in Table 5.2.2. All the combinations of the variables with the regression algorithms performed better than the baseline model. Set 1 and Set 2 with both Random Forest and XGB Regressor showed better performance on the unseen test data. But these two combinations of models also showed higher tendencies of overfitting on the training data. With respect to Radom Forest, Set 1 has a training RMSE of 1.01 and Set 2 has a training RMSE of 1.10. For XGB Regressor, training RMSE of Set 1 is 1.17 and Set 2 is 1.29. It is difficult to pick out best performing model as the other models have similar performance score with roughly about 10% better performance when compared to the baseline model.

Table 5.1.2. Test RMSE of Survey Method

Set Number	Linear Regression	KNN	Random Forest	XGB
Set 1	1.65	1.65	1.55	1.53
Set 2	1.66	1.63	1.55	1.53
Set 3	1.66	1.66	1.64	1.65
Set 4	1.66	1.65	1.64	1.64

Reducing the max_depth (maximum depth a tree can have) parameter to 11 reduces the problem of overfitting for Random Forest with Set 1. The corresponding Training and Test RMSE are 1.33 and 1.55 respectively. By reducing the max_depth to 10, overfitting problem is also reduced in Set 2 for Random Forest. The training RMSE is 1.40 and the Test RMSE is 1.56, when using the new hyperparameter. Similarly, reducing the max_depth in XGB Regressor for Set 1 and Set 2 to 5 and 6 respectively also helped with the problem of overfitting on the training data. For Set 1 the training

RMSE is now 1.47 and the Test RMSE on unseen data is 1.54 and for Set 2 the training RMSE is 1.42 and the test RMSE is 1.53. With the new performance results, XGB regressor with Set 2 variables is the better performing model. This model performs 15% better than the baseline model.

When analyzing the residuals further, using Shapiro-Wilk test from module ‘scipy.stats’ shows the residuals are not normally distributed. The residuals seem to be skewed distribution. An example of the density plot is shown in Fig 5.2.2. The residuals plot shows the residuals plotted against the predicted value. This plot shows that the model predicts lower scores more than higher scores. This could be because larger number of lower scores were available to train on than higher scores of procrastination. Fig 5.2.3. also shows the pattern of not capturing all the variance in modelling procrastination. This is inferred with the linear pattern between residuals and the predicted values that is seen in the residual plot. This usually indicates a temporal dependency between the residuals or presence of important, unmodeled, grouping structure in the data.

Fig 5.2.2. Density plot of residuals of XGB Set 2

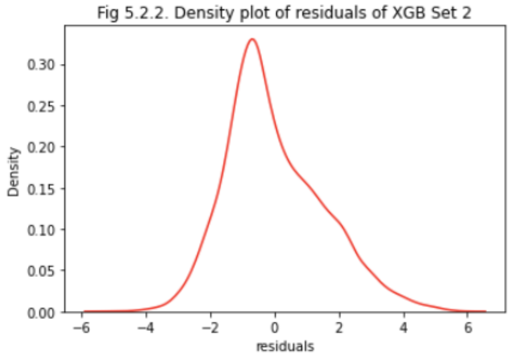
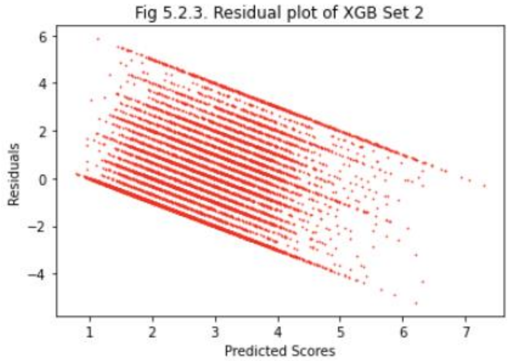


Fig 5.2.3. Residual plot of XGB Set 2



While looking into the important features contributing to the tree-based models such as Random Forest and XGB regressor, Set 2 and Set 3 are used to understand the models better. Set 1 shows a similar pattern to Set 2. Categorical variables such as Day of the week and Type of survey contributed nearly nothing to model the variance. Fig 5.2.4 and Fig 5.2.5. notes the important features contributing to predicting momentary procrastination in Random Forest method.

Fig 5.2.4. Important Features of RF Set 2

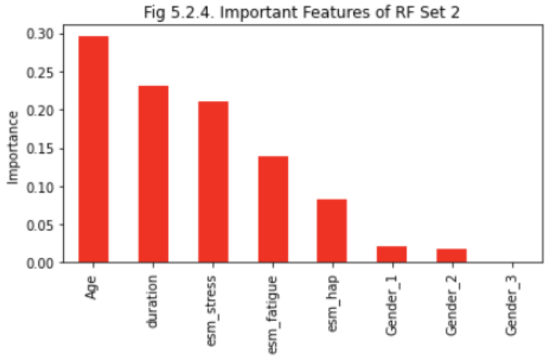
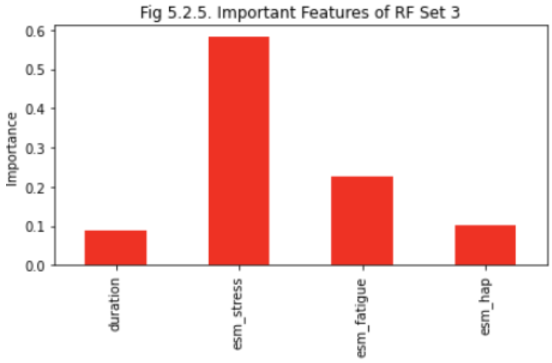
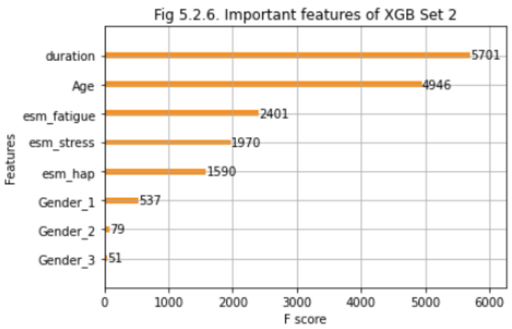
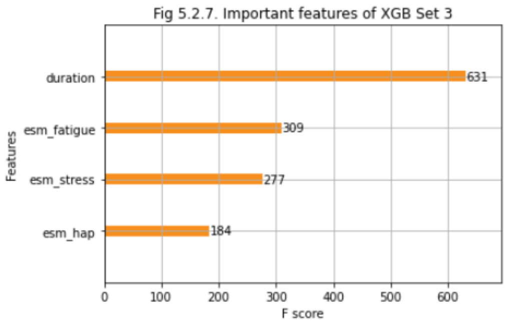


Fig 5.2.5. Important Features of RF Set 3



When Age is included in the mixture, it is the biggest contributor in predicting procrastination and duration is the second biggest predictor. When age is removed, Stress is the biggest predictor and duration is the least contributor in prediction. This pattern is not noticed with the same sets of features in XGB regressor. Also as indicated in literature, XGB Regressor works well in including categorical features better than Random Forest. Fig 5.2.6. and Fig 5.2.7. shows the important features of predicting procrastination when using XGB Regressor.





5.3. Mixed Method :

The final method used in the study is the mixed models where features from Survey Method and ESM are used in modelling. As the dependent variable used in this method is ESM's procrastination score, the baseline RMSE is the same as ESM with **baseline RMSE = 1.805**.

Table 5.2.1. shows the optimal hyperparameters tuned using the training data for each set of variables when using K-Nearest Neighbors, Random Forest, and Extreme Boosting Regressor.

Table 5.2.1. Optimal Hyperparameters

Set No.	KNN	Random Forest	XGB
Set 1	17	max_depth = 15, n_estimators = 300	max_depth = 8, learning_rate = 0.1
Set 2	16	max_depth = 15, n_estimators = 300	max_depth = 8, learning_rate = 0.1
Set 3	19	max_depth = 13, n_estimators = 300	max_depth = 7, learning_rate = 0.09
Set 4	18	max_depth = 14, n_estimators = 300	max_depth = 7, learning_rate = 0.1

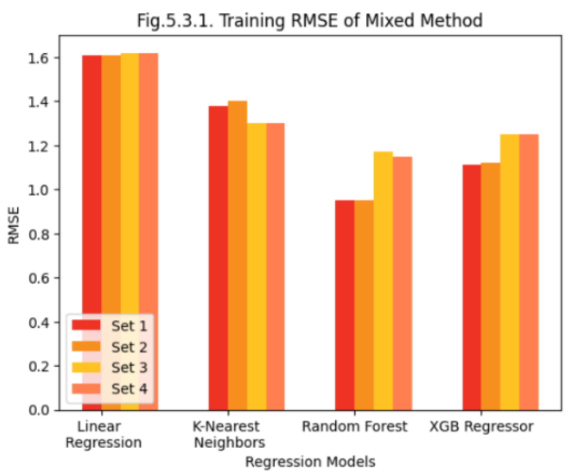


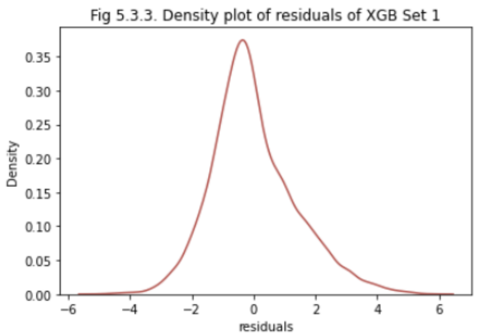
Table 5.3.2. shows the performance of trained models on the out of sample data. By increasing the complexity of the models (adding variables from the Survey method), the results have improved for some algorithms. Except for XGB Regressor with Set 2, 3 and 4, all the other models performed better when compared to the baseline model. Random Forest and XGB Regressor again shows overfitting even with the increase in the complexity of the model. Training RMSE of Random Forest on an average has 0.4-unit difference when compared to the Test RMSE. Whereas XGB Regressor on Set 2,

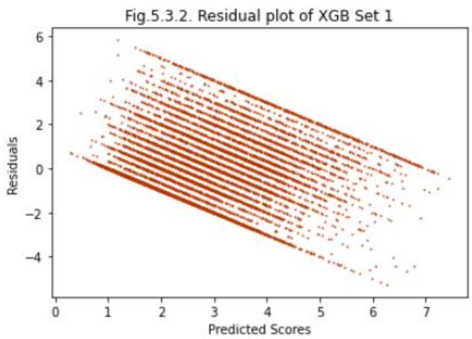
3 and 4 shows a difference of 1-unit between their Training RMSE and Test RMSE. With the current sets of models used in predicting procrastination, KNN with Set 3 has the better performance and the least overfitting with a training RMSE of 1.30 and test RMSE of 1.39. This model performs 23% better when compared to the baseline model.

Table 5.3.2. Test RMSE of Mixed Method

Set Number	Linear Regression	KNN	Random Forest	XGB
Set 1	1.61	1.54	1.39	1.41
Set 2	1.62	1.61	1.41	2.14
Set 3	1.62	**1.39**	1.44	2.15
Set 4	1.63	1.43	1.42	2.18

To fix the overfitting problem occurring in Random Forest and in XGB Regressor, first the depth of each tree allowed in each regression model is decreased. This reduced the problem to some extent in Random Forest but did not help much in XGB Regressor. After reducing the max_depth to 6 for Set 1 in XGB Regressor and adding ‘subsampling = 0.5’ as one of its parameters, the model’s performance improved with Set 1 alone. The training RMSE for this model is 1.26 and the test RMSE of unseen data is 1.38 making it the better model.





Looking further into the residuals, they are probably not normally distributed (Fig 5.3.3). The pattern observed in Fig.5.3.2. which is similar to ESM method residual plots, shows that the residuals are neither random nor independent. These are similar to other residuals plots, irrespective of the algorithm used. This might indicate a temporal dependence or clusters present within the dataset. Increasing the complexity of the model by adding more features, does not seem to help in fixing the residuals of these models.

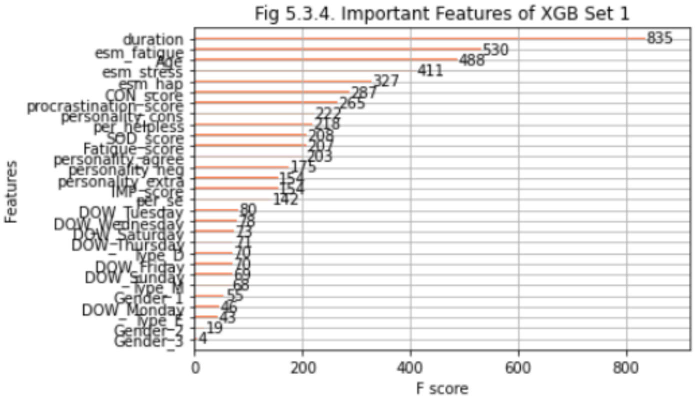
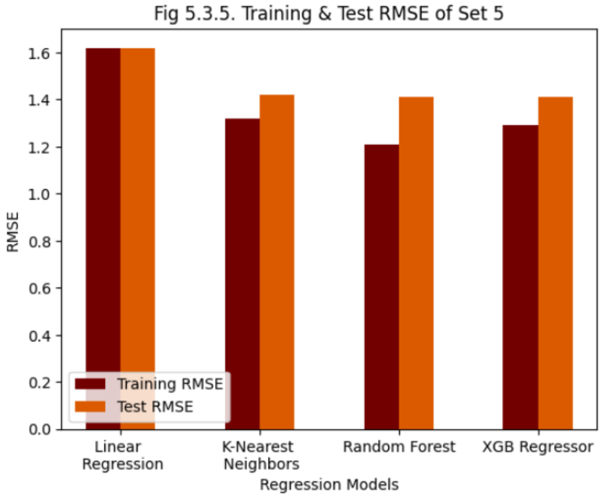


Fig 5.3.4. Important Features of XGB Set 1

The most important features of XGB Regressor model with Set 1 variables are shown in Fig 5.3.4. Looking at this plot, we can see resemblance with the important features from both Survey Method and ESM. Duration feature plays an important role in estimating the procrastination scores followed by ESM variables and then other trait variables. Categorical variables' contribution is minor in both Random Forest as well XGB Regressor. Even when looking into linear regression and KNN models, these models perform better even without these categorical variables. Removing these variables along with other intercorrelated and less contributing variables, Set 5 is created to see if the performance is enhanced with these selective set of variables.

Fig 5.3.5. Training & Test RMSE of Set 5



The feature included in Set 5 are 'Duration', 'Age', 'ESM Stress', 'ESM Fatigue', 'ESM Happiness', 'SM Procrastination', 'SM Conscientiousness' (personality), 'SM Neuroticism' (personality), 'SM Perceived Helplessness', 'SM Fatigue', 'SM Connectedness', 'SM Social Desirability'.

ESM stands for variables collected from Experience Sampling Method; SM stands for variables collected from Survey Method.

Table 5.3.3. Optimal Hyperparameters of Set 5

Model	Hyperparameters
KNN	K = 21
Random Forest	max_depth = 12, n_estimators = 300
XGB Regressor	Max_depth = 7, learning_rate = 0.05

Fig 5.3.5. shows the performance of Set 5 with each algorithm. XGB regressor performs the best with Test RMSE of 1.41 and Training RMSE of 1.29. With subsampling at 0.5, overfitting is reduced. Even though, the difference in performance between KNN, Random Forest and XGB Regressor are negligible, XGB Regressor is chosen as it is the quick to train and robust in performance. The hyperparameters used for each algorithm is mentioned in Table 5.3.3.

6. Discussion

In this section we will talk in detail about the results and how they answer the research questions. The findings of this study are divided among four sub questions. After discussing these findings, they are brought together to answer the main question. We will delve deeper into the impact and limitations of this study and what could further be done in future research.

6.1. Survey Method :

The first sub-question is regarding the Survey Method, "*To what extent variables from the survey method predict trait procrastination ?"*. To answer this, we employed four regression algorithms namely Linear Regression, K-Nearest Neighbors, Random Forest and Extreme Boosting Regressor. The dataset was also divided into four different combinations of features to investigate which combination provides the best result.

Most of the algorithms except for Linear Regression with Set 2 and XGB Regressor with Set 1 and 3, performed better than the baseline model (dummy regressor). Random Forest Regressor model with Set 4 features where Age and Gender is excluded and the personality and stress scores dimension are expanded performed better than the others. This model performed 14% better than the baseline model on unseen data. What does it mean to perform 14% better than the baseline model? When training Random Forest algorithm with these features, this model performs 14% better in predicting unseen data (test data) than when compared to just predicting the mean of the training target scores.

The variables used in the survey method show how far trait procrastination can be predicted. Perceived helplessness which is a facet of perceived stress scale is the biggest contributor in predicting procrastination followed by personality trait Conscientiousness. These features are in line with the literature (Schouwenburg & Lay, 1995; Tice & Baumeister, 1997; Gropel & Steel 2008)). For example, a person scoring high on Conscientiousness could be so diligent in finding the best way to do a task that they spend time researching about the task rather than doing the task. This they could see as procrastination. It is important to note that literature suggested more of a linear relationship which is present but a simple learning algorithm such as Linear Regression could not encapsulate the variance of procrastination described by these features. Since tree-based models don't produce co-efficient like linear models, it is difficult to jot down the precise relationship. This is not a problem as this does not affect the predictive validity of the model.

Things that call for concern are the RMSE scores of the models which is ranging from 0.87 to 1.08. They are quite high even on the best predicting model (RMSE = 0.87). Root mean squared error is unforgiving to high errors and that is exactly what we have here. Looking at the residual plot, the points are scattered further away from 0. The test RMSE then suggests that when predicting procrastination with this model, the margin of error is approximately .87 units for a scale of values between 1 to 5. This is not terrible per se. But there is definitely a margin for improvement.

The reason for high RMSE scores, could be because the model is missing key information or there isn't enough data to train the model. Out of the 235 data points, only 75% is used in training and validating the model (176 instances). The rest 25% is left out for testing (59 instances). This could also be the reason why the performance does not improve when using more robust models such as XGB Regressor.

It is also important to note that predicting human psychological variables is a tough task. When dealing with one-off questionnaire such as the one used in survey method, the validity of the instrument is important (Bell & Gosnell, 2020). Though this study uses validated surveys, humans are heterogeneous and human nature tend to bias the survey responses. This could lead to systematic bias in the dataset that when used in machine learning models could lead to erroneous predictions.

6.2. Experience Sampling Method :

Moving on to the ESM method, the question to be answered is, '*To what extent variables from the ESM predict momentary procrastination ?*'. Similar modelling was done with four different algorithms and four different combination of features. These included psychological features, demographic variables, and the circumstances under which the ESM questionnaires were filled. For example, the duration took to fill the questionnaire, the day of the week the questionnaire was filled and the type of the questionnaire.

Initial modelling with hyperparameters chosen from GridSearchCV showed all the combinations of algorithms with different sets of features to perform better than baseline model. Although the performances were better than baseline model (dummy regressor), nonlinear models like Random Forest and XGB Regressor showed overfitting on training data. Reducing the depth of the individual tree parameter fixed the problem of overfitting. Running the models with the new parameters then produced the best performing model. This was the XGB Regressor with Set 2 features. These features included Age, Duration, Gender, Stress, Fatigue and Happiness. This model performed 15% better than the baseline model. That is, including these features improved prediction of momentary procrastination by 15% when compared to just predicting training target's mean.

While trying to model the variables from ESM, it is important to note that main contributors are the momentary stress, momentary fatigue, and momentary happiness, albeit the final 2 to a lesser degree. The interesting contributors are age and the duration one takes to fill these momentary questionnaires. Studies such as Haycock et al. (1998) and Ferrrari & Diaz-Morales (2007) showed no relationship between age and procrastination whereas in this study, age plays a major contributory role in predicting momentary procrastination with nonlinear models (Random Forest and XGB

Regressor). At the first glance, the effect of age and duration seems to be exaggerated since removing age diminishes the contribution of duration in Random Forest algorithm with Set 3. It could be that the model is learning random noise but XGB Regressor also shows similar contributions with Age and Duration. Removing these variables and using only psychological variables seem to make the learners simple and thereby resulting in high bias and low variance (Briscoe & Feldman, 2011). This makes the model stable but low on predictions.

The variables extracted from ESM seems to be hardly adequate in modelling momentary procrastination. The residuals paint the same picture. The residuals show a non-normal distribution. A non-normal distribution of target variables could lead to non-normal residuals (Pek et al., 2018). The pattern noticed in the residual plots shows multiple linear data points. Two potential inferences could be taken from this. One, the model is missing important predictor variable maybe a variable that is not even measured in this dataset. This could be the reason for high RMSE scores. Two, it could suggest a serial dependency between the residuals. There is still no strong enough evidence to say concretely that either of those issue is present. Also, boosting algorithm such as the XGB Regressor should learn from the residuals of previous trees and improve its predictions. Looking the residual plot of XGB regressor is similar to residual plots of other algorithms. This means that the models are learning the same wrong information suggesting a high bias.

To summarize, variables from ESM show potential in predicting momentary procrastination. It uses interesting combinations of variables with non-linear models. However, there seems to be a absence of important variables in predicting momentary procrastination. To check if this could be fixed with adding other possible predictor variables, the mixed method is used next.

6.3. Mixed Method :

Analyses from the Mixed Method answer the following question, "To what extent variables from the Survey Method and ESM predict momentary *procrastination?*". A similar pattern is followed with four regression algorithms and four combinations of the features set in modelling predictive models. ESM procrastination is used as the target variable as we are trying to see if trait behavior could add important variance in predict momentary behavior.

When adding the variables from survey method with ESM variables, the performance of distance-based algorithm as well as non-linear algorithms are better. Compared to the baseline models, KNN with Set 3 performed 23% better. This non-parametric method found an approximation of associations between the dependent and independent variables that is modelled without overfitting on the training data. After reducing the depth of trees in Random Forest and XGB Regressor, overfitting on the training data is reduced and XGB Regressor with Set 1 to perform equally as good as KNN model. This model with KNN uses lesser variables which makes it efficient. Still, XGB Regressor processes data quicker than KNN when processing larger datasets, this makes it a better model. XGB Regressor with Set 2, 3 and 4 performance did not improve even after trying to fix the problem of overfitting. This algorithm shows high variance when not using right combination of features in mixed methods. This high variance makes these models unstable.

Adding complexity to the model does not change much with the residuals. They are still not normally distributed with similar parallel lines pattern occurring in the residual plot. Along with temporal dependency, there seems to be also indications of clustering occurring in these plots (Long & Travedi, 1993). There are indications of sub-groups within the dataset that the models are missing while modelling the dataset. This is plausible since the momentary experiences are from different participants collected over a period of 30 days that is grouped together.

It is important to mention that residuals of predicting ESM procrastination seem to violate the assumptions of Linear Regression analysis. This will be of concern if we were trying to explain and interpret the relationships between these variables. As the main objective of this study is about predicting future instances, violating these assumptions does not make these models unstable (Rosenbusch et al., 2020).

Going back to the overarching question about the comparison of variables from the different method of data collection, it is important to ask which type of procrastination is to be predicted. Different methods of data collection pertain to different types of variables to be predicted. Survey method relates to trait variables whereas ESM relates to measuring state (momentary) variables. Depending upon the task at hand these need to be defined.

For example, when trying to find an ideal job candidate, trait procrastination is needed to be predicted effectively. It is then important to have a model that could effectively predict procrastination using as less information as possible. This is where the costs and benefits come to play. It is quite expensive to obtain the scales used in this modelling. Also, the prediction error is approximately 0.87. this means that when the model predicts a candidate to have a 3 on procrastination scale, this prediction needs to be taken as “3 +/- 0.87” that is the difference between a low to average to above average procrastinator. This is quite a discrepancy. Also, the context needs to be considered.

Independent analysis on this dataset has shown correlation between trait procrastination and average momentary procrastination. It means that participants who had scored high on trait procrastination (onboarding survey), had higher average momentary procrastination scores and vice versa. However, there is stronger correlation between average momentary stress and average momentary procrastination. That is participants who had experienced higher stress levels while reporting momentary stress also reported higher momentary procrastination levels.

This association is also seen in the models used in this study. ESM stress is one of the big contributors in predicting ESM procrastination. However, to model the variance of momentary procrastination effectively, trait variables are required. Experience Sampling Method of data collection can be useful in diagnostics to understand how one's situation affects their procrastination level (Wessel et al., 2020). As assessments are made in the natural environment, there exist high ecological validity in Experience Sampling Method of data collection (Aalbers et al., 2020). Still models used in this study to predict momentary procrastination are very expensive.

To be able to predict momentary procrastination, the models developed in this study uses 13 variables (Set 5 from Mixed Models). This uses variables from ESM and Survey Method. This would also require developing an application to administer the questionnaire on a regular basis. From a practical standpoint, these self-report surveys could be administered once daily as a reflection of the whole day or once in the end of the week as a reflection of the whole week. This constant self-reflection could be beneficial, but it is also quite a tedious process. After data collection, maintaining this database safe is crucial as this involves sensitive information. This is expensive.

Even though it might be expensive to establish experience sampling methodology on a large scale to help identify individuals at risk of exhibiting extreme procrastination, it might be helpful on the longer run when considering the facts that procrastination leads to reduction in quality of work (Ferrari, 2001), wasting potential when not engaging them effectively (Metin, Peeters and Taris, 2018) or overworking them into burnout (Hall et al., 2019). Not identifying at risk individuals is more expensive in the longer run.

From a prediction standpoint, it is less expensive to collect data using survey method and it also more straightforward to use them for prediction. By increasing the data points, these analyses should be repeated to check its predictive performance. Predicting momentary experiences ask for more complex models which is difficult to collect and model. Hence, simple is better when predicting procrastination. Overall, this study gives a baseline for future studies to use when trying to predict procrastination with psychological variables.

6.4. Limitations and Future research :

This study has some limitations. More than 10,000 entries were omitted due to incomplete surveys in the experience sampling responses. Future research could find ways to use these data to have better predictions. Even with using almost all the features, the features recorded on the monthly basis were not used for this study. This may add to better model fit in predicting procrastination. Also, the participant pool used in this study is not representative of the general population. These are young adults in their bachelor program with average age of 20.5 years. Out of 235 participants in this study, only 16 participants were above the age 25 years. Hence, generalizing these predictions to the general working-class population must be done with caution.

Also, a major part of the ESM questionnaires was filled during the first Covid-19 lockdown. The weak correlation in ESM method could be because of the type of dataset that is collected. For example, participants even when they felt they had procrastinated in that time frame, could have felt less stressed about it because there was suddenly a surplus in time available to complete tasks. This also explains why there is a positive correlation between fatigue and happiness. Participants were trying to find different ways to keep themselves engaged to be happy (not feel lonely or stressed about the pandemic), that they felt mentally fatigued in trying to do so. Future research could divide the data into responses before the Covid-19 lockdown and responses during Covid-19 lockdown and analyze if the results differ significantly. Another interesting idea for future study could be to check with this dataset is if participants who scored higher on social desirability scale on an average had lower scores measuring less desirable personality and behavior traits. Also, the residuals suggested different clusters within the dataset suggesting partial temporal dependency. Grouping the responses at the participant person level to see if this indeed shows a temporal dependency would be a great starting point for future studies.

7. Conclusion

The goal of this study is to see to what extent variables collected from different methods of data collection predict procrastination. This is done through assessing them separately and then together to find the optimal model in predicting procrastination. The costs involved in collecting these data and the benefits it could produce is discussed.

This study offers unique contributions about, to what extent state and trait procrastination could be predicted using machine learning models. Trait procrastination was predicted 14% better than the baseline and momentary procrastination was predicted 23% better than the baseline. Trait procrastination required lesser features. On the other hand, the predicting momentary procrastination delivered better RMSE scores but required complex featuring to get higher predictive scores. This proved to be overall expensive. Depending on the task at hand, benefits of investing in predicting these variables is subjective.

However, this study provides a good baseline for future studies to use when trying to predict procrastination using machine learning algorithms and psychological predictors.

References

Aalbers, G., Keijsers, L., Vanden Abeele, M. M. P., & Hendrickson, A. T. (2020, October 19). Preregistration of data collection. Retrieved from osf.io/6fs92

Aalbers, G., Abeele, M. V., Hendrickson, A. T., De Marez, L., & Keijsers, L. (2021). Caught in the moment: Are there person-specific associations between momentary procrastination and passively logged smartphone use?

Abbasi, M., Dargahi, S., Pirani, Z., & Bonyadi, F. (2015). Role of procrastination and motivational self-regulation in predicting students' academic engagement. *Iranian journal of medical education*, 15, 160-169.

Akram, A., Fu, C., Li, Y., Javed, M. Y., Lin, R., Jiang, Y., & Tang, Y. (2019). Predicting students' academic procrastination in blended learning course using homework submission data. *IEEE Access*, 7, 102487-102498.

Altman, N. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3), 175-185.

Bell, M., & Gosnell, S., (2020). Using self-report survey data to predict human psychology is tough. https://www.mercuryds.com/blog/using-self-report-survey-data-to-predict-human-psychology-is-tough

Bhattacherjee, A. (2012). Social science research: Principles, methods, and practices.

Briscoe, E., & Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition*, 118(1), 2-16.

Brownlee, J. (2021). XGBoost for Regression. https://machinelearningmastery.com/xgboost-for-regression/#:~:text=Extreme%20Gradient%20Boosting%20(XGBoost)%20is,of%20the%20gradient%20boosting%20algorithm.&text=XGBoost%20is%20an%20efficient%20implementation,used%20for%20regression%20predictive%20modelling.

Chauhan, R. S., MacDougall, A. E., Buckley, M. R., Howe, D. C., Crisostomo, M. E., & Zeni, T. (2020). Better late than early? Reviewing procrastination in organizations. *Management Research Review*.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY, USA: ACM. https://doi.org/10.1145/2939672.2939785

Choi, J.N. and Moran, S.V. (2009), “Why not procrastinate? Development and validation of a new active procrastination scale”, *The Journal of Social Psychology*, Vol. 149 No. 2, pp. 195-211.

D’Abate, C. P., & Eddy, E. R. (2007). Engaging in personal business on the job: Extending the presenteeism construct. *Human Resource Development Quarterly*, 18, 361–383. doi:10.1002/hrdq.1209

Das, M. (2012). Innovation in online data collection for scientific research: The Dutch MESS project. *Methodological Innovations Online*, 7(1), 7-24.

Daumé III, H. (2012). A course in machine learning. *Publisher, ciml. info*, 5, 69.

Davydov, D. (2014). What is Procrastination Costing Your Business? https://www.bitrix24.com/about/blogs/business-development/what-is-procrastination-costing-your-business-.php

Drăgulescu, B., Bucos, M., & Vasiu, R. (2015). Predicting assignment submissions in a multi-class classification problem. *TEM Journal*, 4(3), 244.

Ferrari, J. R., & Ware, C. B. (1992). Academic procrastination: personality. *Journal of social Behavior and personality*, 7(3), 495-502.

Ferrari, J. R., Johnson, J. L., & McCown, W. G. (1995). Procrastination and task avoidance: Theory, research, and treatment. *Springer Science & Business Media*.

Ferrari, J. R., O'Callaghan, J., & Newbegin, I. (2005). Prevalence of procrastination in the United States, United Kingdom, and Australia: arousal and avoidance delays among adults. *North American Journal of Psychology*, 7(1).

Ferrari, J. R., Diaz-Morales, J. F., O'Callaghan, J., Diaz, K., & Argumedo, D. (2007). Frequent behavioral delay tendencies by adults: International prevalence rates of chronic procrastination. *Journal of Cross-Cultural Psychology*, 38(4), 458-464.

Gadosey, C. K., Schnettler, T., Scheunemann, A., Fries, S., & Grunschel, C. (2021). The intraindividual co-occurrence of anxiety and hope in procrastination episodes during exam preparations: An experience sampling study. *Learning and Individual Differences*, 88, 102013.

Glen, S. (2019). Decision Tree vs Random Forest vs Gradient Boosting Machines: Explained Simply. https://www.datasciencecentral.com/profiles/blogs/decision-tree-vs-random-forest-vs-boosted-trees-explained#:~:text=Like%20random%20forests%2C%20gradient%20boosting,one%20tree%20at%20a%20time.

Gustavson, D. E. (2016). *Investigating the cognitive underpinnings of procrastination: An intervention study and a longitudinal analysis* (Doctoral dissertation, Doctoral Dissertation). University of Colorado, Denver, The United States).

Gröpel, P., & Steel, P. (2008). A mega-trial investigation of goal setting, interest enhancement, and energy on procrastination. *Personality and individual differences*, 45(5), 406-411.

Hall, N. C., Lee, S. Y., & Rahimi, S. (2019). Self-efficacy, procrastination, and burnout in post-secondary faculty: An international longitudinal analysis. *PloS one*, 14(12), e0226716.

Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature 585, 357–362 (2020).

Haycock, L. A., McCarthy, P., & Skay, C. L. (1998). Procrastination in college students: The role of self-efficacy and anxiety. *Journal of counseling & development*, 76(3), 317-324.

Hong, J.-C., Lee, Y.-F., & Ye, J.-H. (2021). Procrastination predicts online self-regulated learning and online learning ineffectiveness during the coronavirus lockdown. *Personality and Individual Differences*, 174, Article 110673.

Hooshyar, D., Pedaste, M., & Yang, Y. (2020). Mining educational data to predict students' performance through procrastination behavior. *Entropy*, 22(1), 12.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90–95.

James, G., Witten, D., Hastie, T., Tibshirani, R., Hastie, M. T., & Suggests, M. A. S. S. (2017). Package 'ISLR'. *R-Project*. October, 19.

Kármen, D., Kinga, S., Edit, M., Susana, F., Kinga, K. J., & Réka, J. (2015). Associations between academic performance, academic attitudes, and procrastination in a sample of undergraduate students attending different educational forms. *Procedia-Social and Behavioral Sciences*, 187, 45-49.

Kpotufe, S. (2011). k-NN regression adapts to local intrinsic dimension. *arXiv preprint arXiv:1110.4300*.

Kumcagiz, H., Ersanli, E., & Alakus, K. (2014). Hopelessness, procrastination and burnout in predicting job satisfaction: A reality among public school teachers. *International Journal of Academic Research*, 6(1), 333-339.

Lay, C. H. (1990). Working to schedule on personal projects: An assessment of person-project characteristics and trait procrastination. *Journal of Social Behavior and Personality*, 5(3), 91-103.

Lay, C. H. (1992). Trait Procrastination and the Perception of. *Journal of Social Behavior and Personality*, 7(3), 483-494.

Lay, C. H., & Schouwenburg, H. C. (1993). Trait Procrastination, Time Management. *Journal of social Behavior and personality*, 8(4), 647-662.

Lay, C. H. (1997). Explaining lower-order traits through higher-order factors: the case of trait procrastination, conscientiousness, and the specificity dilemma. *European Journal of Personality*, 11(4), 267-278.

Long, J. S., & Trivedi, P. K. (1993). Some specification tests for the linear regression model. *Sociological Methods & Research*, 21(2), 161–204. doi: 10.1177/0049124192021002003

Malachowski, D., & Simonini, J. (2006). Wasted time at work still costing companies billions in 2006. Retrieved from http://www.salary.com/wasted-time-at-work-still-costing-companies-billions-in-2006/

McCown, W., Petzel, T., & Rupert, P. (1987). An experimental study of some hypothesized behaviors and personality variables of college student procrastinators. *Personality and Individual Differences*, 8(6), 781-786.

McKinney, W., & others. (2010). Data structures for statistical computing in python. *In Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).

McNeish, D., & Matta, T. (2018). Differentiating between mixed-effects and latent-curve approaches to growth modelling. *Behavior Research Methods*, 50(4), 1398-1414.

Metin, U. B., Peeters, M. C., & Taris, T. W. (2018). Correlates of procrastination and performance at work: The role of having "good fit". *Journal of prevention & intervention in the community*, 46(3), 228-244.

Milgram, S. (1992). McGraw-Hill series in social psychology.The individual in a social world: Essays and experiments *(2nd ed.)* (J. Sabini & M. Silver, Eds.). *Mcgraw-Hill Book Company*.

Olive, D. M., Huynh, D. Q., Reynolds, M., Dougiamas, M., & Wiese, D. (2019). A quest for a one-size-fits-all neural network: Early prediction of students at risk in online courses. *IEEE Transactions on Learning Technologies*, 12(2), 171-183.

Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.

Pek, J., Wong, O., & Wong, A. (2018). How to address non-normality: A taxonomy of approaches, reviewed, and illustrated. *Frontiers in psychology*, 9, 2104.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Rajapakshe, W. (2021). The Impact of Academic Procrastination, Self-Efficacy, and Motivation on Academic Performance: Among Undergraduates in Non-State Universities in Sri Lanka. *Asian Journal of Education and Social Studies, 25-38*.

Ramsay, J.O. (2006), “Functional data analysis”, *Encyclopedia of Statistical Sciences,* Vol. 4, pp. 1-9.

Refaeilzadeh, P., Tang, L. & Liu, H. (2009). Cross-Validation.. In L. Liu & M. T. Özsu (ed.), *Encyclopedia of Database Systems (pp. 532-538)* . Springer US . ISBN: 978-0-387-39940-9.

Reinecke, L., Meier, A., Beutel, M. E., Schemer, C., Stark, B., Wölfling, K., & Müller, K. W. (2018). The relationship between trait procrastination, internet use, and psychological functioning: Results from a community sample of German adolescents. *Frontiers in psychology*, 9, 913.

Rosenbusch, H., Soldner, F., Evans, A. M., & Zeelenberg, M. (2019). Supervised machine learning methods in psychology: A practical introduction with annotated R code. Social and Personality Psychology Compass, e12579.

Sirois, F. M., & Giguère, B. (2018). Giving in when feeling less good: Procrastination, action control, and social temptations. *British Journal of Social Psychology*, 57(2), 404-427.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5), 1763-1768.

Schouwenburg, H. C., & Lay, C. H. (1995). Trait procrastination and the big-five factors of personality. *Personality and Individual differences*, 18(4), 481-490.

Schouwenburg, H. C. (2004). Procrastination in Academic Settings: General Introduction.

Shmueli, G. (2010). To explain or to predict?. *Statistical science*, 25(3), 289-310.

Sirois, F. M. (2014). Procrastination and stress: Exploring the role of self-compassion. *Self and Identity*, 13(2), 128-145.

Steel, P. (2007). The Nature of Procrastination: A Meta-Analytic and Theoretical Review of Quintessential Self-Regulatory Failure. *Psychological Bulletin*, 133(1), 65-94. https://doi.org/10.1037/0033-2909.133.1.65

Tice, D. M., & Baumeister, R. F. (1997). Longitudinal Study of Procrastination, Performance, Stress, and Health. *Self-Regulation and Self-Control*, 299-309.

Varghese, D. (2018). Comparative Study on Classic Machine Learning Algorithms, Part -2, https://medium.com/@dannymvarghese/comparative-study-on-classic-machine-learning-algorithms-part-2-5ab58b683ec0

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Ostblom, J., & Lukauskas, S. (2018, 07). mwaskom/seaborn: v0.9.0. Retrieved from https://doi.org/10.5281/zenodo.1313201

Wessel, J., Bradley, G. L., & Hood, M. (2019). Comparing effects of active and passive procrastination: A field study of behavioral delay. *Personality and Individual Differences*, *139*, 152-157.

Wessel, J., Bradley, G. L., & Hood, M. (2020). A low-intensity, high-frequency intervention to reduce procrastination. *Applied Psychology*.

Xu, Y. (2019, June). Proud and Productive Procrastination?. In *International Conference on Digital Transformation and Global Society* (pp. 661-671). Springer, Cham.

Yang, Y., Zha, K., Chen, Y. C., Wang, H., & Katabi, D. (2021). Delving into Deep Imbalanced Regression. *arXiv preprint arXiv:2102.09554*.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100-1122.

Zhang, S., & Feng, T. (2020). Modelling procrastination: Asymmetric decisions to act between the present and the future. *Journal of Experimental Psychology: General*, 149(2), 311.

Appendix

Appendix A.

```
['EthicaID','DeviceID','ScheduledTime','IssuedTime','ResponseTime','Dur
ation','Location','S1','S2','S3','F1','F2','F3','P1','P2','P3','H1','Ty
pe','Age','Sex','Gender','BDD1','BDD2','BDD3','BDD4','BDD5','BDD6','BDD
7','BDD8','BDD9','BDD10','BDD11','BUR1','BUR2','BUR3','BUR4','BUR5','BU
R6','BUR7','BUR8','BUR9','BUR10','BUR11','BUR12','BUR13','BUR14','BUR15
','BUR16','MDD1','MDD2','MDD3','MDD4','MDD5','MDD6','MDD7','MDD8','MDD9
','MDD10','MDD11','MDD12','MDD13','MDD14','MDD15','MDD16','MDD17','MDD1
8','MDD19','MDD20','MDD21','FAT1','FAT2','FAT3','FAT4','FAT5','FAT6','F
AT7','FAT8','FAT9','FAT10','FAT11','FAT12','FAT13','FAT14','FAT15','FAT
16','FAT17','FAT18','FAT19','FAT20','PSS1','PSS2','PSS3','PSS4','PSS5',
'PSS6','PSS7','PSS8','PSS9','PSS10','PRO1','PRO2','PRO3','PRO4','PRO5',
'CON1','CON2','CON3','CON4','CON5','CON6','CON7','CON8','BFT1','BFT2',
'BFT3','BFT4','BFT5','BFT6','BFT7','BFT8','BFT9','BFT10','BFT11','BFT12
','BFT13','BFT14','BFT15','BFT16','BFT17','BFT18','BFT19','BFT20','BFT2
1','BFT22','BFT23','BFT24','BFT25','BFT26','BFT27','BFT28','BFT29','BFT
30','IMP1','IMP2','IMP3','IMP4','IMP5','IMP6','IMP7','IMP8','MOR1','MOR
2','MOR3','MOR4','MOR5','SOD1','SOD2','SOD3','SOD4','SOD5']
```

Key:

S - Stress (Experience Sampling Method)	F - Fatigue (Experience Sampling Method)
P - Procrastination (Experience Sampling Method)	H - Happiness (Experience Sampling Method)
BDD - Body Dysmorphic Disorder	BUR - Burnout
MDD - Major Depressive Disorder	PSS - Perceived Stress Scale
Pro - Procrastination	CON - Connectedness
BFT - Big Five Personality Item	IMP - Impulsivity
MOR - Morning-ness/Evening-ness	SOD - Social Desirability

Each number refers to number of the question per questionnaire as their responses were noted separately.

Appendix B

Density plots

ESM features :

Figure 1 Fatigue

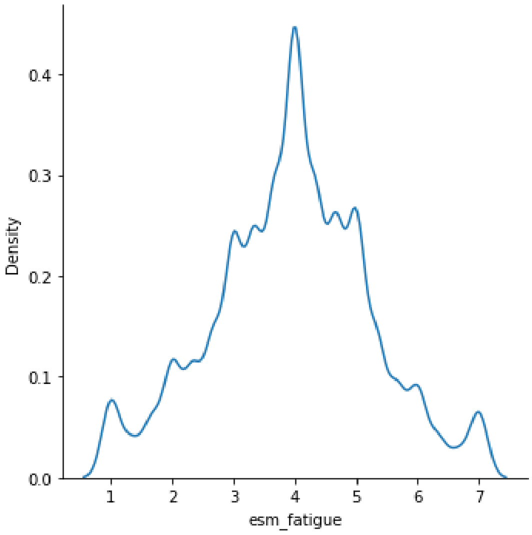


Figure 2 Happiness

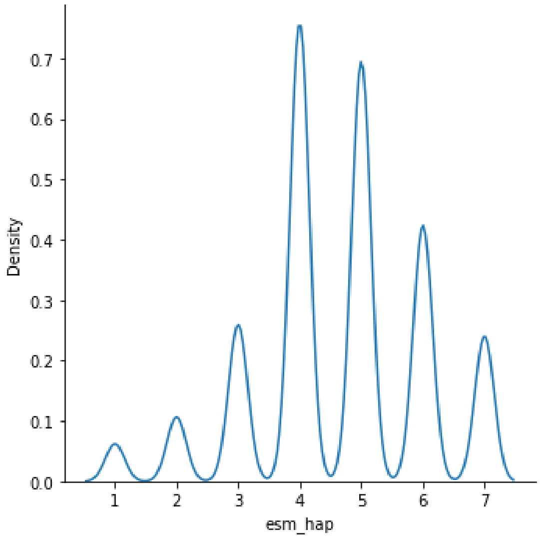


Figure 3 Procrastination

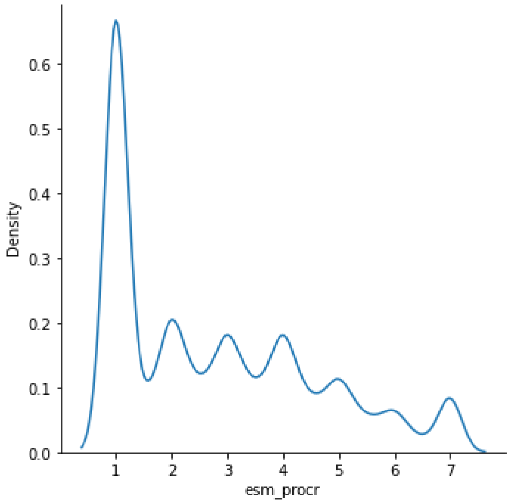
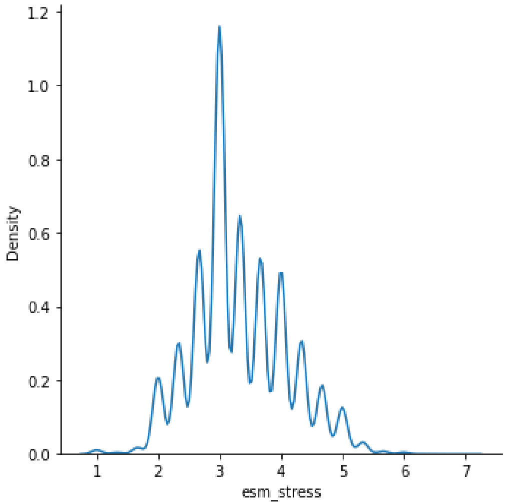


Figure 4 Stress



Survey Method features :

Figure 5 Agreeableness

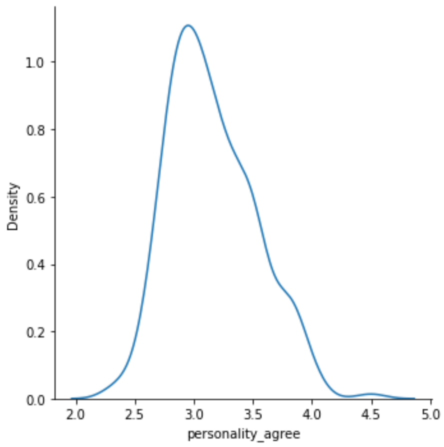


Figure 6 Social Connectedness

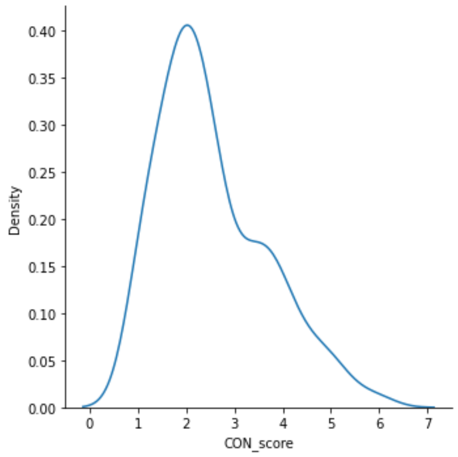


Figure 7 Conscientiousness

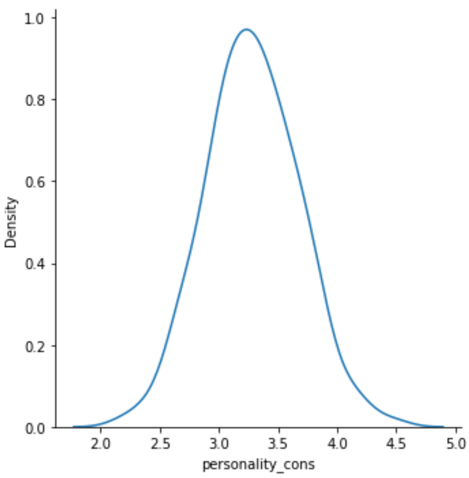


Figure 8 Extraversion

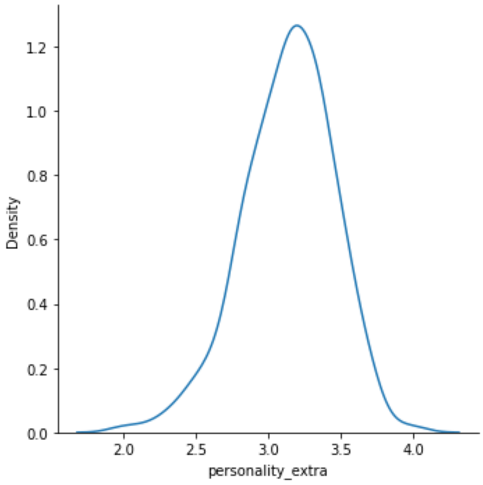


Figure 9 Conscientiousness

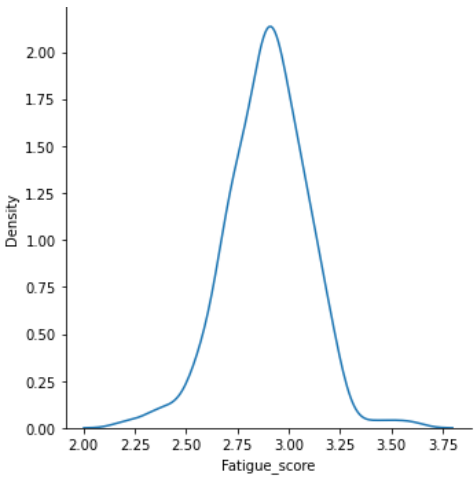


Figure 10 Helplessness

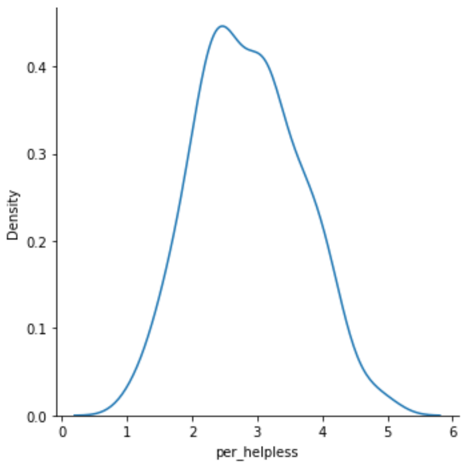


Figure 11 Procrastination

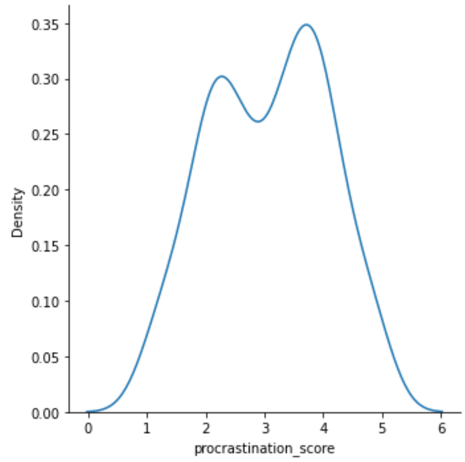


Figure 12 Impulsivity

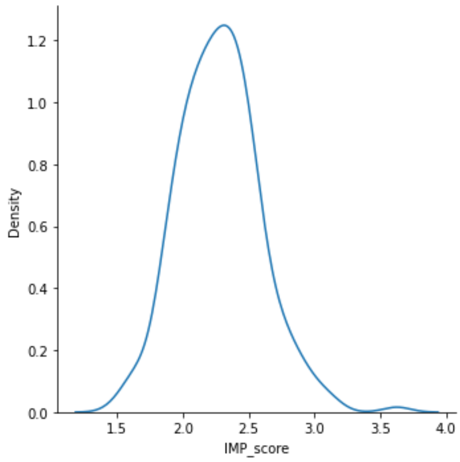


Figure 13 Openness

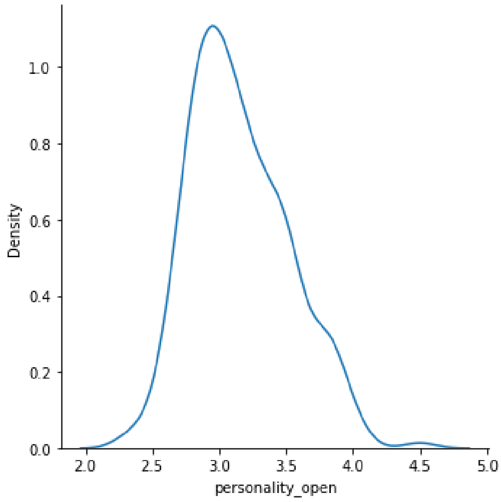


Figure 14 Neuroticism

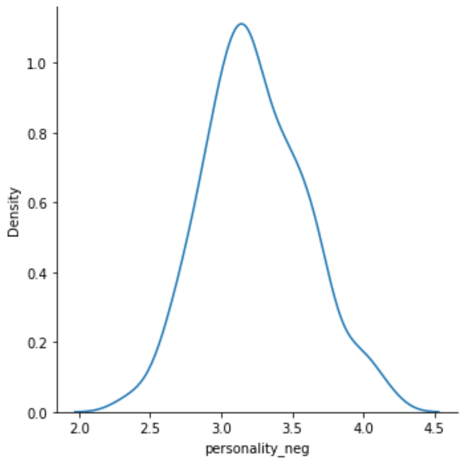


Figure 15 Self-Efficacy

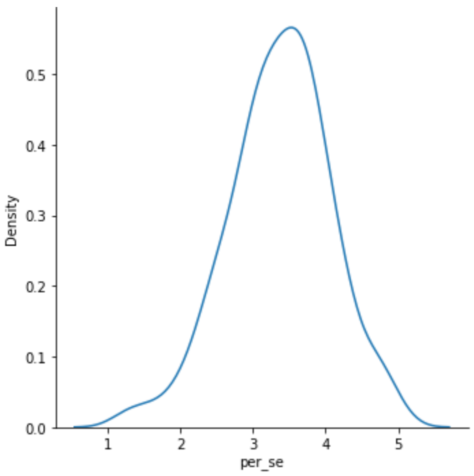


Figure 16 Social Desirability

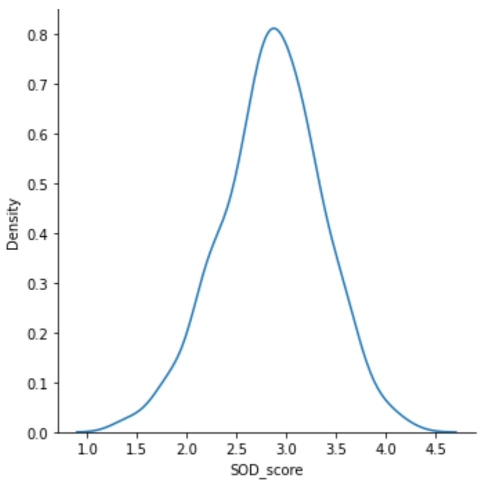


Figure 17 Stress

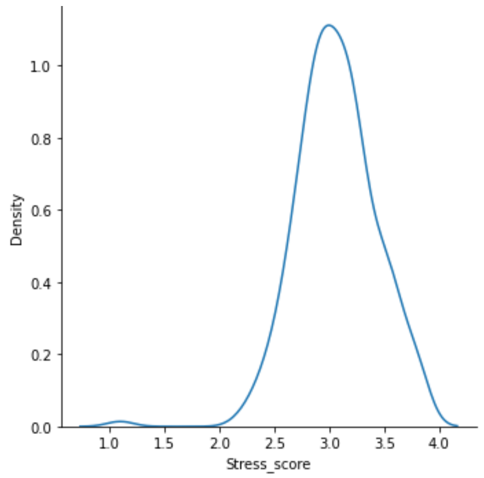


Figure 18 Personality