Predicting Personality Traits Using Smartphone Sensor Data and App Usage Data

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

in Partial Fulfillment of the Requirements

for the degree of Master of Science

in the Department of Computer Science

University of Saskatchewan

Saskatoon

By

Naveen Kumar Kambham

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

    Head of the Department of Computer Science

    176 Thorvaldson Building

    110 Science Place

    University of Saskatchewan

    Saskatoon, Saskatchewan

    Canada

    S7N 5C9

ABSTRACT

Human behavior is complex – often defying explanation using traditional mathematical models. To simplify modeling, researchers often create intermediate psychological models to capture aspects of human behavior. These intermediate forms, such as those gleaned from personality inventories, are typically validated using standard survey instruments, and often correlate with behavior. Typically these constructs are used to predict stylized aspects of behavior. Novel sensing systems have made tracking behavior possible with unprecedented fidelity, posing the question as to whether the inverse process is possible: that is, inferring psychological constructs for individuals from behavioral data. Modern smartphones contain an array of sensors which can be filtered, combined, and analyzed to provide abstract measures of human behavior. Being able to extract a personal profile or personality type from data directly obtainable from a mobile phone without participant interaction could have applications for marketing or for initiating social or health interventions. In this work, we attempt to model a particularly salient and well-established personality inventory, the Big Five framework [48]. Daily routines of participants were measured from parameters readily available from smartphones and supervised machine learning was used to create a model from that data. Cross validation-based evaluation demonstrated that the root mean squared error was sufficiently small to make actionable predictions about a person's personality from smartphone logs, but the model performed poorly for personality outliers.

ACKNOWLEDGEMENTS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EDS	Ethica Data Systems
RMSE	Root mean Squared Error
MSE	Mean Squared Error
CV	Cross Validation
ReLU	Rectified Linear Unit
RBF	Radial Basis Function
ROC	Receiver Operating Characteristic
SLP	Single Layer Perceptron
MLP	Multilayer Perceptron
RNN	Recurrent neural network
CNN	Convolutional neural network
KNN	K-Nearest Neighbors
SVM	Support Vector Machines
MBTI	Myers-Briggs Type Indicator
MMPI	Minnesota Multiphasic Personality Inventory
FFM	Five Factor Models
NEO PI-R	The Revised NEO Personality Inventory

Chapter 1

Introduction

Who we are underpins what we do: this is the central premise of personality inventories used in the social and health sciences [21, 106]. Intermediate representations of personality provide more convenient representations of the potential for behavior, and have seen successful application to problems in business [10, 60, 104], health [6, 119, 121] and psychology [26, 27, 32].

Understanding and determining human personality can help many sectors of society, for example, human resources departments make use of personality tests to assess candidates' fit for a job and tests can be used to form a balanced team that can improve the overall performance as a team [11, 71]. In the health sector [52, 79, 118, 121], establishing personality can help in diagnosing the mental health or other contributing factors associated with a range of other symptoms. Identifying peoples' behavior helps businesses design items based on personal interests [76, 94]. Personalities were used in tailoring the marketing materials to prospective customers [70, 95, 109]. Recently, technology designers have investigated the automatic adaptation of games to suit differing personalities [95].

Typically, establishing/understanding the personality is done through the use of questionnaires, such as the Big Five Personality Test [48] or the Myers-Briggs inventory [93]. In questionnaire-based methods, participants are asked to report about themselves by answering questions such as "is talkative?", "remains calm in tense situations?", "how many friends did you talk to?", etc. Based on the responses, personality traits are estimated. Among several personality frameworks, the Big-Five Personality Test is among the most frequently used in psychology, though it has not gained global recognition or been treated as a standard. Several researchers have critiqued the Big Five framework for limiting the personality to only five dimensions and its inability to capture other important behavioral characteristics such as honesty or humor [49, 99]. In spite of this, the Big Five framework continues to be one of the most widely employed personality inventories and shows consistent results across different populations [10, 47, 66, 67, 96, 111, 127]

The Big Five framework consists of five psychological features, namely *Openness*, *Conscientiousness*, *Extroversion*, *Agreeableness*, and *Neuroticism*. A person's personality is a mix of all five traits. For example, an individual personality derived from the test is *Openness* 40%, *Extroversion* 60% , *Agreeableness* 55%, *Conscientiousness* 80%, and *Neuroticism* 45%. In this case, the dominant trait is *Conscientiousness*, but the individual also exhibit traits related to *Extroversion* and *Agreeableness*, and less so *Neuroticism* and *Openness*. A brief summary of traits based on the descriptions in [48] is provided below.

Openness: Openness refers to how open one's mind is to new ideas and situations requiring flexibility. It reflects the adventurous nature of people and their earnest desire to know or learn new things. People who are open tend to be more receptive to new suggestions and influences.

Conscientiousness: Conscientious people are disciplined and have an organized lifestyle. They tend to do things in an orderly and timely fashion. People with high conscientiousness work efficiently due to their systematic thought process.

Extraversion: An extrovert is a person who is socially adept and enjoys being with other people. Interpersonal interactions make such people feel energized and are therefore an essential part of their lives. They are opportunity seekers and think positively.

Agreeableness: This behavioral characteristic is associated with the cooperativeness of a person. People with this trait are usually in compliance with others and are not argumentative. They work well in team situations because they are socially harmonious and show empathy for others.

Neuroticism: Negative-minded people are usually characterized by this trait. They are often victims of anger, jealousy, and depression. Upon encountering a problematic situation, they easily feel threatened and become anxious. They can suffer from an inferiority complex, do not trust people, and are most likely to be stressed.

Table 1.1 shows the personality characteristics for high scorers and low scorers of the corresponding Big Five personality trait.

Trait	Low Scorers	High Scorers
Openness	Conventional, Close-Minded, Routine	Creative, Curious, Exploratory
Conscientiousness	Lazy, Disorganized, Negligent	Organized, Consistent, Hard-Working
Extraversion	Passive, Reserved, Dull	Talkative, Outgoing, Energetic
Agreeableness	Suspicious, Critical, Unwelcoming	Friendly, Cooperative, Generous
Neuroticism	Calm, Composed, Stable	Sensitive, Nervous, Worrying

Table 1.1: Big Five Personality Traits with low and high scorers. Table reproduced based on [16]

Several studies have shown that self-reported personality traits (of any test) are temporally limited and do not capture potential variability in personality which changes over time [13, 42, 126]. Questionnaires can be limited as they demand direct interaction to complete the survey. Self-report procedures have many well-known biases because of memory limitations, fluctuation in mood, and limited participant awareness of critical behaviors [50, 98, 126]. Other techniques for understanding the behaviors are laboratory studies by presenting hypothetical scenarios to participants. Though these methods are better than questionnaire-based methods, they do not capture the behavioral aspects of humans leading normal lives. These limitations have motivated research into new, low-effort, and reliable ways of collecting data about individual behavior, which might be linked to personality.

Smartphones have become a part of everyday life in developed societies. Smartphones allow individuals to

stay in touch with one another and support many day-to-day tasks, including navigation, Internet browsing, online shopping, or attending virtual meetings. There are over 2.6 billion smartphone users worldwide [116] with 87% of Millennials saying that their phones are always by their side [88]. This makes smartphones a compelling platform to collect the behavioral data in natural lives for the studies of human personality. Contemporary smartphones are host to numerous physical and logical sensors related to location, communication, phone state (e.g. screen lit, charging status), phone orientation, and interaction (connection to other devices or the Internet). These sensors can be used to derive activities like movement, interaction, and daily habits [4,61]. Human spatial behavior constructs have been predicted by psychological constructs, for example, *conscientious* people (a psychological construct derived from the Big Five test) are more likely to have, and stick to, a schedule [48]. *Extroverts* will spend more time in groups, in contact with others, and recording these experiences [48]. Knowing that these psychological constructs correlate with physically measured behaviors, it is logical to ask if psychological constructs could be predicted from human spatial behavior, or more specifically, can spatial and communications behaviors observable via smartphone sensor and app usage data be used to predict personality?

Individual differences can be explored and established with data collected from smartphones [57]. It is the purpose of this study to establish whether such differences extend to those associated with validated personality traits. For example, users who are open to using different applications and make use of the camera and maps might exhibit the Big Five *Openness* trait more than those who do not. Similarly, users who spend more time on messaging applications like Viber, Whatsapp, and SMS may score higher for *Extraversion*. *Conscientiousness* might be represented by following the same daily routine (time of arriving/leaving university, battery charging time, etc.). These are hypothetical and anecdotal examples that serve as guidelines to study the correlation of data with personality traits [24,53,108]

Smartphones as a platform for data collection may allow us to observe the behavior of humans in a much more natural way than in a potentially contrived laboratory setting. Exploring this data in the context of personality traits might even enable us to see changes in these traits over time (as expressed through changes in behavior). More importantly, it is an opportunity to study different situations in daily life that might be affected by personality. For example, are conscientious people less susceptible to weather disruptions in the daily routine? Feeding back such patterns of behavior could in turn be used as the basis for interventions to benefit individuals.

My primary hypothesis was that real-world behavior, measured with sufficient precision from Smartphone sensor data, could be used to predict personality. Relationships between different input features extracted from smartphone data and the self reported Big Five personality traits (*Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism*) was analyzed through machine learning models. Unlike most prior work, personality is modeled as a regression problem, placing participants on a continuum, rather than in discrete bins. Regression modeling was chosen over classification modeling because the Big Five Personality Inventory is a continuous representation of personality.
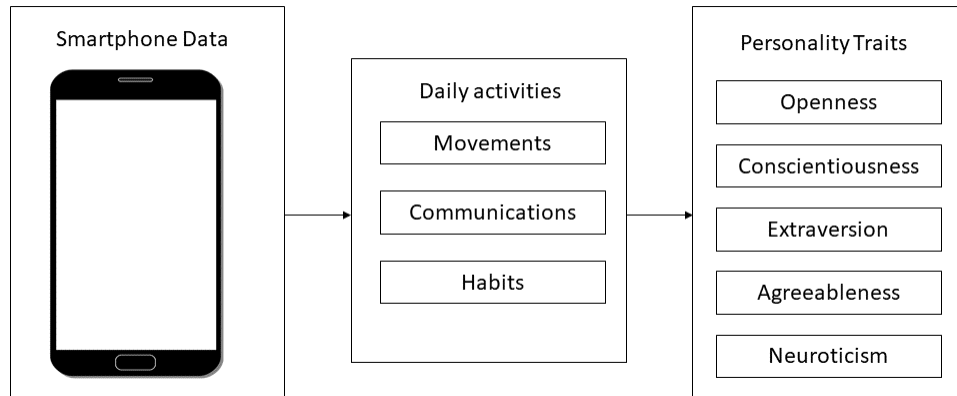
Figure 1.1: Graphical Representation of the Hypothesis

Results are substantially better than all other regression models and most classification models reported in the literature. Unlike prior work, results are compared to a baseline model: continually guessing the population mean. While the results were superior to those reported in the literature, they were comparable to guessing the population mean in terms of statistical metric comparison. A closer look at the fit line plots and residual plots indicated that models provide more informative predictions than baseline models, as the error in the baseline model is systemic, not random. This work provides both a contribution in improving state of the art in inferring personality from measurement and also provides a more viable comparator algorithm which should be employed in future research on this topic.

1.1 Organization of Thesis

This thesis is comprised of seven chapters. Chapter 2 provides the background information relevant to this thesis work. Existing research work pertinent to human behavior using various forms of technology-mediated data and their limitations are discussed in Chapter 3. Chapter 4 presents the experimental setup for collecting the required data and also detailed explanations about feature extractions. Chapter 5 describes the methodologies and machine learning models employed. Chapter 6 presents the results and finally, Chapter 7 summarizes the results and major contributions of this work.

Chapter 2

Background

2.1   Supervised Machine Learning

In supervised learning, a model is provided with sets of input and output pairs, known as training data, and the model employs some form of optimization to attempt to find a set of parameters which minimize the error to produce the desired output for a given input. Models are trained according to the optimization criterion usually until a target accuracy (specified by a measure of fitness) is reached or a maximum number of training iterations have been performed. There are two major types of supervised machine learning models:

- Regression Models

- Classification Models

Regression Models

If the output value or the value to be predicted is a continuous value then regression models are used to learn from the data. For example, predicting the Big Five Personality traits of an individual is a regression modeling problem if the personality trait values are real numbers.

Classification Models

In classification, the model objective is to predict a class label from all possible classes. For example, classifying whether an email is a spam or not is a classification problem. A regression problem can also be posed as a classification problem by binning the continuous values.

Three supervised machine learning models, namely Neural Networks, Random Forests, and Support Vector Machines were used in this work. A summary of the three machine learning models is presented in the following sections, neural networks are presented in detail as the personality traits were modeled using neural networks and the rest were used for comparing the quality of the machine learning models created.

2.1.1   Neural Networks

Artificial neural networks are approximation algorithms inspired from the architecture of biological neurons. Signals are propagated and mathematically fused through a series of non-linear basis calculations and linear
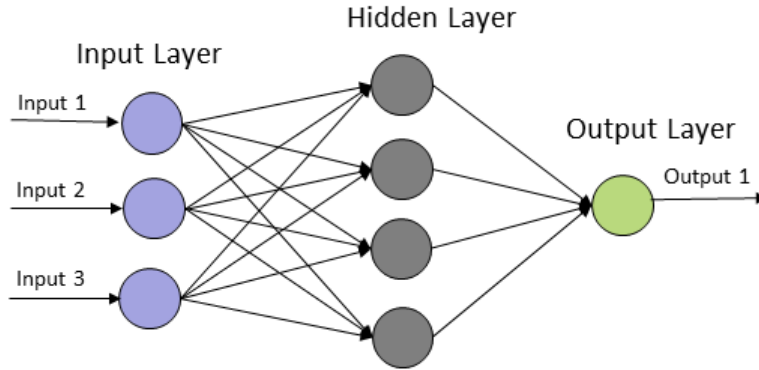
Figure 2.1: Sample Neural Network created for demonstration purpose

sums. There are several variants of Neural Network models: Single Layer Perceptron (SLP), Multilayer Perceptron (MLP), Recurrent neural network (RNN), and Convolutional neural network (CNN). For an overview of the various neural network architectures and their applications, see [54, 62, 112].

Many classic neural networks architectures contain three layers: an input layer, a hidden layer, and an output layer. The input layer represents the independent variables of the model. The output layer represents the dependent variables (values to be predicted) in the model. The hidden layers represent an approximation of the mapping between the input and output layers and enable the network to learn the complex relationships between independent variables and output variables. Each layer in the network consists of several neurons. The input layer contains the neurons representing the independent variables of the model and, similarly, the output layer contains the neurons representing the dependent variables (values to be predicted) in the model. In the case of the hidden layer, the number of neurons depends on the complexity of the data to be learned. Neurons are typically connected to all other neurons in the next layer, via a weighted connection. Within each neuron, an activation function expresses how the weighted sum of inputs should be treated and, in the hidden layer, introduce non-linearity into the models. This creates a network of simple neurons enabling highly complex transformations between the layers. Figure 2.1 shows a classic three layer neural network, Multilayer Perceptron (MLP). Here, the neural network model is created to predict one output feature based on three input features. The input layer contains three independent features namely *Input1, Input2,* and *Input3*. The hidden layer contains four neurons, and the output layer contains only one neuron *Output1*.

Training a neural network consists of two steps. Initially, all connections (synapses) are assigned random weights. In the forward pass, the hidden neuron values, and output neuron values are calculated using the activation functions and the input values, as shown in equation 2.1, equation 2.2 and equation 2.3. Based on the expected output and predicted output, error or loss of the model is calculated using a loss function, typically a variant of the squared error shown in equation 2.4. In the backward pass, an optimizer is used to reduce the loss/error by adjusting the weights associated with each neuron as shown in equation 2.6. This whole process is repeated until the network model produces results closer to the expected output. Such

6

networks were successfully used in various problems such as pattern recognition (speech recognition, blood cell classification) [51, 59, 86] and prediction (stock market prediction, weather forecasting) [7, 69, 72, 105]. Neural networks have been used for decades to provide models of complex systems when and where more traditional statistical techniques fail [12, 62, 74]

$$net_{h1} = w_1 \times input1 + w_2 \times input2 + w_3 \times input3 \tag{2.1}$$

where: w_1, w_2, w_3 are weights in the Hidden layer

$$out_{h1} = Activation\ Function(net_{h1}) \tag{2.2}$$

$$\widehat{output1} = w_{11} \times out_{h1} + w_{12} \times out_{h2} + w_{13} \times out_{h3} + w_{14} \times out_{h4} \tag{2.3}$$

where: $w_{11}, w_{12}, w_{13}, w_{14}$ are weights in the Output layer

$$error = RMSE(output1, \widehat{output1}) \tag{2.4}$$

where: $\widehat{output1} = predicted\ output1$

$$error = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (output1_i - \widehat{output1}_i)^2} \tag{2.5}$$

$$\frac{\partial(error)}{\partial w_{11}} = \frac{\partial}{\partial \widehat{output1}} \left(\sqrt{\frac{1}{n} \sum_{i=1}^{n} (output1_i - \widehat{output1}_i)^2} \right) \times \frac{\partial \widehat{output1}}{\partial w_{11}} \tag{2.6}$$

2.1.2   Random Forests

Random Forests are an ensemble learning method for decision trees. A decision tree is a graph structure created by splitting the data repeatedly into subsets, usually according to a single feature. A classic decision tree learning algorithm creates a model by splitting along the input dimension of greatest variance according to a heuristic or a cost function. Essentially, decision trees learn a hierarchy of (often true/false binary) decisions, leading to a classification of the data. A typical problem with decision tree learning is its tendency to overfit the data, that is to model the noise in the data as well as the underlying trend in the data itself, leading to poor model generalizability. To avoid these overfitting problems, random forests are employed. Random forests, as shown in Figure 2.2, divide the whole data set into random small subsets (without replacement) and the decision tree is constructed for each subset. An aggregate statistic (usually the mean or mode) of the output of the ensemble (forest) of decision trees is taken as the actual answer. For an overview of the various decision tree architectures, learning modes and applications see [17].

7
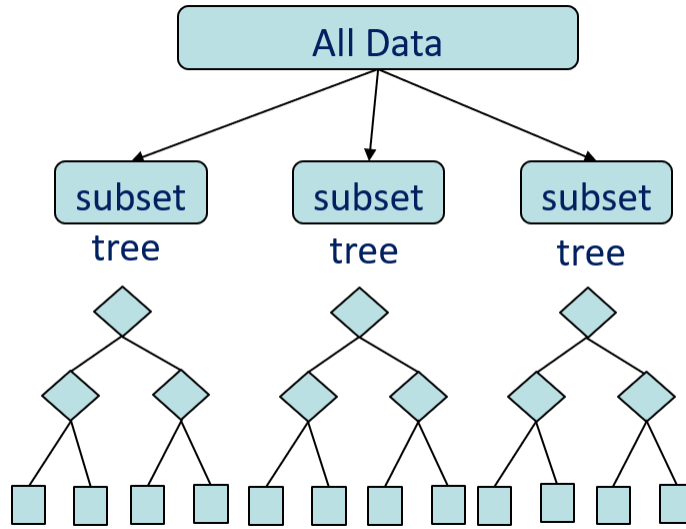
Figure 2.2: Sample Random Forests model created with three decision trees for the purpose of demonstration, image reproduced based on [84]

2.1.3 Support Vector Machine

A Support Vector Machine (SVM) is a supervised classification machine learning model. In SVM, the model attempts to find a hyperplane that linearly separates the data into corresponding classes. Hyperplanes can be a simple line or a higher dimensional plane depending on the data set. Several hyperplanes are possible for a dataset. The best hyperplane maximizes the margin between the classes. Real data often cannot be linearly separated as shown in Figure 2.3. In such cases, the feature space can be enlarged and transformed into higher dimensions by adding new features. Special functions called *Kernels* are used to quantify and add the features based on similarities. This whole process of enlarging feature space can help the model to generalize the non-linearity in the data. SVMs have been successfully used for various applications such as text classification, spam detection, and image classification [23, 34, 64, 117].
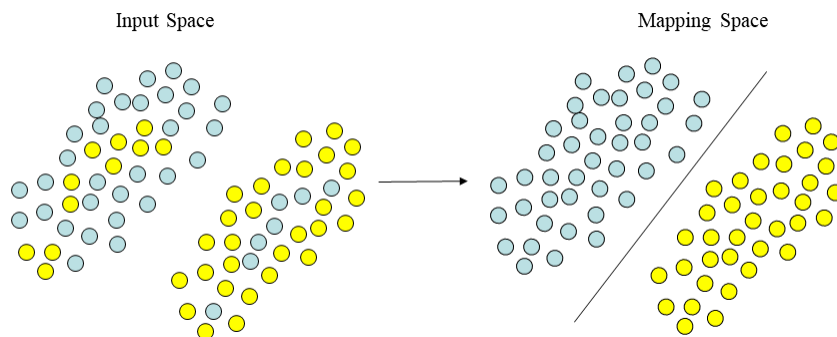
Figure 2.3: Visual representation of Support Vector Machine transforming the Non-liner separable data in to higher dimensional space

2.1.4 Baseline Models

Baseline models also called random models are simple statistical models which use summary statistics to model the phenomenon in question. Baseline models are used to compare the quality of machine learning models. Baseline models can be created for both regression and classification modeling problems. In this work, two baseline models namely the Mean model, and ZeroR model were used.

Mean Model

A Mean model is a baseline model for regression modeling problems. In a mean model, the output value is predicted as the population mean irrespective of the input values. For example, Table 2.1 shows five participants' data with three input features and an output feature (actual trait). The predicted trait column represents the mean model predicted values. As shown, the mean model returned 0.6 as the predicted trait for all the participants because the mean of the population is 0.6.

Participants	Input 1	Input 2	Input 3	Actual Trait	Predicted Trait
Participant 1	15	5	7	0.5	0.6
Participant 2	5	7	8	0.2	0.6
Participant 3	8	10	17	0.8	0.6
Participant 4	21	16	5	0.5	0.6
Participant 5	7	12	14	1	0.6

Table 2.1: Sample Data for Mean Model

ZeroR Model

ZeroR model is a baseline model for classification modeling problems. In ZeroR, the model returns the most frequent class label of the population as the predicted class label irrespective of the input features. For example, Table 2.2 shows five participants' data with three input features and output class labels represented by the actual trait column. Here possible output class labels are high (H) or low (L). Since the majority of the population has 'H' as the class label, ZeroR model returns 'H' as predicted label for all participants.

Participants	Input 1	Input 2	Input 3	Actual Trait	Predicted Trait
Participant 1	15	5	7	H	H
Participant 2	5	7	8	H	H
Participant 3	8	10	17	H	H
Participant 4	21	16	5	L	H
Participant 5	7	12	14	L	H

Table 2.2: Sample Data for ZeroR Model

2.2   Model Evaluation

2.2.1   Cross-validation

Cross-validation is a technique to evaluate or check the consistency of machine learning models. In cross-validation, the entire data set is divided into K folds from which one set is kept aside as a validation set and the model is trained using the remaining K-1 sets of data. This method allows for testing against unseen data without potentially biasing the results by randomly choosing a favorable or unfavorable test set. For example, in a ten-fold cross-validation, the entire data is divided into ten sets. In Round 1, one set is kept aside as a validation set, and the model is trained using the nine sets. The same process is repeated for remaining sets. The accuracy or error for each round is calculated separately, and the average of all ten rounds is taken as the final accuracy or error of the model.

2.2.2   Root Mean Squared Error

The root mean squared error is a quality metric for regression models. It is computed by finding the square root of the mean of the squares of the difference between the actual values and predicted values. Since this is an error metric, a model with lower root mean squared error is considered as a better model. For the sample mean model data shown in Table 2.1 RMSE value is 0.28.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (true_i - predicted_i)^2}$$

2.2.3   Accuracy

Accuracy is a statistical quality metric for classification models. It is computed as a ratio of the number of correct predictions made by the model to the total number of tests. A model with high accuracy is considered better. For the sample classification data shown in Table 2.2, the model made three correct predictions out of five tests. Hence the accuracy of the model is 60%.

Chapter 3

Literature Review

3.1 Traditional Methods

Traditional methods for understanding human personality are participant observation and survey-based methods. In participant observation methods, participants' daily routines/habits are collected through observation for a period of time or by reading diaries participants' are asked to keep. Based on these observations/notes participants' personalities are estimated. Though these methods provide reliable and accurate estimates for the personality, they are difficult, expensive, and time-consuming to conduct on a large population. These limitations paved the way for self-reporting methods known as questionnaire/survey-based methods. Popular survey-based methods are Minnesota Multiphasic Personality Inventory (MMPI) [28], Myers-Briggs Type Indicator (MBTI) [93], NEO personality inventory [27], and Big Five Personality Inventory [48]. In survey-based methods, participants answer behavioral questions. These responses are used to estimate the personality traits.

Several types of personality tests are available depending on the requirements and situations. For example, MMPI is a popular personality test for clinical research while MBTI and Big Five tests are widely used for employee assessments. Initially, personality inventories were very long: MMPI consists of 570 item questions and takes a considerable amount of time to finish. Researchers focused on developing simple Five Factor Models (FFM) [32]. In FFM models personality is divided into five dimensions or traits. Among several five-factor models, the Big-Five personality test is frequently used in psychology.

Research has shown that self-reported personality inventories have many biases because of memory limitations, fluctuation in mood and limited participant awareness of their critical behaviors [13, 42, 50, 98, 126]. Other traditional techniques for understanding the personalities are laboratory studies, in which participants are presented with hypothetical scenarios. Though these methods are better than questionnaire-based methods, they are necessarily artificial. These limitations have motivated research into new, low-effort, and reliable ways of collecting data about human behavior. Existing research has dealt with human personality prediction using various forms of technologically mediated data collection. Some researchers used Internet usage, and others used social networking activities or smartphone data to infer the personality.

3.2   Internet

The rise of the Internet in the 90's made it an inexpensive platform for researchers to conduct behavioral studies. Several researchers experimented with and suggested web-based questionnaires as an alternative to paper-based questionnaires because they are less expensive and easily reach a wide range of people. Web-based questionnaires simplified the generalization of studies by allowing heterogeneous samples and allowed researchers to easily maintain the collected electronic data [18,29,97]. Despite these notable advantages, web-based questionnaires are prone to the problems suffered by self-reporting methods as the only improvement it offered was the medium of data collection, but the core methodologies remained the same.

Researchers explored individuals' Internet usage and established the correlations between Internet usage and personality traits [37,55,58,83,85]. Researchers conducted a large study with 6900 adults' Internet data and found statistical correlations between Big Five traits and Internet usage [83]. Internet usage was collected using a self-reported Internet usage questionnaire on a 6 point scale ranging from using Internet several times a day to 35 times a week. The questionnaire also consisted of questions related to the type of Internet usage, such as: (using the Internet to) send or read email, message friends, play games, work on research for school or job, pay bills, or buy something. Based on the responses, Internet usage was categorized into communication, leisure, academic, or economic. Statistical analysis between the extracted features and self-reported Big Five personality traits indicated that Extraversion, Openness, and Neuroticism were positively related to online communication and Extraversion, Openness, and Conscientiousness were strongly related to leisure activities.

Studies indicated that Internet usage and the personality traits varied based on gender [55]. Internet usage from 72 students was collected and categorized into 12 different Internet services. Neuroticism and Extraversion traits were collected through the Eysenck Personality Inventory [39]. Based on the analysis they found that for men the Extraversion trait was positively associated with the leisure services (random browsing and pornography) and Neuroticism was negatively correlated with the use of information services (work and study info). In the case of women, they found that Neuroticism positively related to use of social services. In contrast to this study other researchers found that found that Agreeableness, Conscientiousness, and Extraversion are negatively correlated with Internet use [75]. They used 117 university students to see relationships between Internet usage and Big Five traits. Participants answered two questionnaires, one for the Big Five test and the other for Internet usage. They collected the number of hours spent on the Internet for different types of purposes: communication, leisure (music, shopping), and academic usage. Surprisingly, few researchers found no relationships with Extraversion or Neuroticism for Internet usage and reported that people tend to use/engage in Internet activities when there is nothing else to do [58]. This is further bolstered by other researchers who found no relationships among 220 participants' self-reported Big Five traits and Internet usage [37].

3.3 Audio, Text, Visual Analysis

Verbal behaviors and linguistic features extracted from written test and audio/video recordings were used to understand the human personalities [53,101,115,123,124]. Participants personalities in a meeting environment were detected through the way they speak and interact in the meetings using audio-visual features [101]. Four participants' meeting video recordings (6 hours duration) were used to extract the features: speech rate, pitch and energy, head orientation, and head, hand, and body fidgeting. Using the extracted features and a supervised learning classification model based on SVM, they successfully predicted the Extraversion trait with 89.1% accuracy. Other researchers demonstrated that it is possible to detect all big five personality traits automatically using audio features and text [82]. They collected 2,479 essays and audio recordings of 96 casual conversations from student participants, along with the Big Five questionnaire. This was all done for ground truth. Using this data, they extracted syntactic and semantic features, such as the ratio of pronouns, questions, assertions, positive emotion words, and anger words. They experimented and modeled the problem using classification, regression, and rank based models and reported that the ranking based models produced the best results with ranking losses of 0.37, 0.33, 0.26, 0.31 and 0.39 for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism respectively. Finally, YouTube videos were successfully used to estimate the Big Five traits [15]. Using 442 bloggers' YouTube videos, visual features and audio features (speaking time, speaking turns, voicing rate, energy, and pitch) were extracted. Stepwise linear regression was used to predict Extraversion trait with 34% variance.

Personal emails were studied to understand the Big Five Personality traits [115]. Over 100,000 email messages were collected from 486 participants' Gmail and Outlook programs. Self-reported Big Five traits were converted to discrete bins of low, medium, and high classes. Emails were used to extract features, namely bag-of-words, meta-features including TO/CC/BCC counts, syntactic labels using part-of-speech tagging, and writing styles (greeting patterns, closing patterns and emojis). SVM was used to predict the personality traits Conscientiousness, Extraversion, Agreeableness, and Neuroticism with an accuracy of 70%. Other researchers conducted a similar large-scale study but with written texts of 2,588 students to predict the Big Five personality traits [124]. Textual features such as word sentiment, speech acts (negotiate, greet, deliver and remind), punctuation (repetition, emojis), bag-of-words, and parts of speech were extracted. Big Five traits were binned into two classes, low and high. SVMs were trained to predict the personality traits Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism with accuracies of 81.9%, 80.4%, 76.8%, 83.36%, 80.2% respectively. They also ran another experiment by binning the personality values into three classes: Low, Medium, High. They achieved an accuracy of around 70%. It can be noticed that making more bins reduced accuracy, as the model needs to find the best support vectors to distinguish three classes instead of two.

3.4   Social Media

Social media provides the opportunity for easy social interactions and self-representation. Social media profiles contain individual information such as personal interests and enable the users to express and share their personal views easily. Millions of these social media profiles made it an alluring platform to understand users [36, 68, 73]. Several researchers have employed online social media data such as participant behavior on social networking sites to predict the users personalities [1, 8, 9, 25, 45, 108].

Researchers successfully demonstrated the correlations between Facebook use and Big Five personality traits [107]. Facebook usage of 97 participants was collected through a 28 item questionnaire, and they noticed substantial correlations between the Big Five personality traits and Facebook use. They made the following observations. Extroverted individuals spend more time on Facebook with more Facebook friends and are part of more Facebook groups than introverts. An individual with high Neuroticism trait tends to reveal private information. Individuals scoring high on Agreeableness have more number of friends. Conscientious people spend less time on Facebook. A similar study was conducted with more diverse participants (1324 including Facebook users and non-users) and different personality tests such as the Big Five inventory, Narcissistic personality inventory, The Revised Cheek and Buss Shyness Scale, and the Social and Emotional Loneliness Scale for Adults Short version [108]. Facebook usage and personality traits were collected using questionnaires. Statistical analysis indicated that Facebook users tend to score high in Extraversion and Neuroticism and non-Facebook users score high in Conscientiousness. These observations are aligned with previous research.

Other researchers used the readily available basic Facebook profile information to identify personality [1,9]. Contrary to the previous self-reporting studies, these studies employed dedicated researchers who looked at the Facebook profiles of participants and collected the required information. Research assistants gathered basic information such as gender, family members, political and religious views, personal interests, favorite movie, and TV shows, contact information, and education and work information [1]. Based on ANCOVA analysis between the collected data and the self-reported personality traits of NEO-PI-R questionnaire they found that Extroverted people have more friends than the others. Other researchers conducted a similar study on 236 Facebook users with diversity in participants (from the US and Germany) and also in personality tests by using both Big Five and the NEO Five-Factor inventory [9]. Nineteen research assistants rated the Facebook profile data by answering the 10-point questionnaire. Based on correlation analysis they demonstrated that Facebook profile data could be used to understand the Openness and Extraversion traits.

Automatic Facebook profile data collection software was used to collect the Facebook information for other behavioral studies [46]. They collected 167 users' personal information, activities, preferences, and language features based on the written descriptions about themselves. Using dedicated multiple linear regression models for each personality trait, they predicted the Big Five personality traits with mean absolute error rates of 11%. A similar study was conducted with a large population of 180,000 using Facebook profile data

collection software myPersonality [8]. From each participant they extracted the number of Facebook friends, number of group associations, Facebook "likes" and "tags" count, number of photos uploaded by users, and other parameters. Using the self-reported Big Five traits and a multivariate linear regression model, they were able to predict Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism with root mean squared errors of 29%, 28%, 27%, 29%, 28% respectively. Higher error rates were attributed to diversity and size of the population. Researchers combined Facebook usage data with the instant messaging data and modeled the Big Five personality test [25]. Facebook and instant messaging usage was collected on a 10 point scale questionnaire. Using self-reported Big Five traits and multivariate analysis they demonstrated that people who scored high on Extraversion and Openness use social media more frequently, and emotionally stable people use social media less.

Several researchers explored Twitter, another popular social media platform, to understand personality [102]. Based on 335 individuals' Twitter accounts they collected: number of followees, number of followers, and number of times a user was listed. Using this information they categorized profiles as: listeners (following many others), popular (followed by many), and highly-read (often listed in reading lists), and influentials (whose tweets were clicked/replied/re-tweeted more often). Extracted features and decision trees were used to predict personality traits with root mean square error around 0.88 on a 1-5 scale. They reported that people who scored high on Openness have more followers and influentials score high on Conscientiousness. Other researchers created a Twitter application that collected the participants' Big Five questionnaire responses and collected Twitter data from 50 users automatically [45]. Based on this data they extracted each individual's number of followers, number of followees, the density of the social network, number of mentions, number of hashtags, links, and number of words per each tweet. Using regression algorithms (Gaussian Process and ZeroR) they could identify Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism with mean absolute errors of 11.9%, 14.6%, 16%, 12.9%, 18.2% respectively.

3.5   Smart Phones and Wearable Sensors

Technological advancements in smartphones have lead to the support of several everyday tasks, these include: navigation, reminders, online shopping, banking, and social interactions. Such advances have lead to smartphones being a part of everyday life in developed societies. To enhance user experience, smartphones are embedded with several sensors and communication interfaces such as GPS, Bluetooth, WiFi, microphone, gyroscope, and accelerometer. These sensors are explicitly intended to aid in positioning, navigation, orientation, and communications. These technological advancements enable opportunities to explore and study various domains such as social networking [87], transportation [89,120], environment [3,92,103], health care [52,80], semantic dimensions of location and movement to describe non-spatial dimensions such as daily routines and behaviors [35,40,77,81,122,122].

Geo-spatial data collected from smartphones was used to identify the places of interest so that recom-

mendations for the new nearby places such as restaurants and shopping centers can be delivered [33,91]. One hundred and thirty-two participants' smartphone sensor data including GPS, WiFi, GSM, and accelerometer data was used to determine places of interest such as home, work, shopping, restaurant, or nightclub [91]. Using grid-based clustering techniques and WiFi routers they successfully extracted stay points for each participant and identified locations for 63% of the day. In a similar study, with 114 participants it was demonstrated that accuracy could be increased to 80% [33]. They found that frequently visited places such as home, work, and friends' home can be easily identified. In both studies, the ground truth was obtained through questionnaires. Cell tower connections were successfully used to understand location-driven behaviors [40]. Thirty students' cellular tower connections from 32,000 cell towers were used to find the patterns in connections. Unsupervised learning method based on hierarchical Bayesian models were used to extract the life patterns. Similar studies were conducted using wearable GPS devices [5,77]. In contrast to these studies, researchers built a behavioral model to predict daily human activities from several smartphone sensors [81]. In their approach, they combined data from several smartphone sensors and used one month of collected data to create a behavioral model. They used the trained model to identify the behavior for the following 20 days with a success rate of 80%.

Smartphone usage has also been used to estimate personality. Researchers collected smartphone usage through self-reports and studied the correlations between personality traits and smartphone usage. One hundred and twelve participants' self-reported smartphone usage data was employed to estimate personality traits [20]. Personality traits were collected using Coopersmith self-esteem inventory and NEO-FFI questionnaires. Participants answered an 8-item smartphone usage questionnaire reporting the weekly time spent on making and receiving calls, sending and receiving SMS, playing games, changing ringtone/wallpaper, and other activities. Based on regression analysis, they reported that Extroverts spend more time on calls and changing wallpapers. Self-reported mobile gaming behaviors were used to identify the personality traits [100]. Regression analysis indicated that people who score low on Agreeableness tend to use mobiles for playing games.

Later, several researchers explored the opportunities to estimate human behavior using automatic extraction of smartphone sensor data, phone call information and app usage data. An overview of various available smartphone sensors and specific areas of psychological research was presented [56]. App usage logs from smartphones were used to predict human personality [113, 125]. In these studies, the app logs were collected and app usage categorized based on the type of usage. Usage was put in the following groups: communications, tools, productivity, games, media, and finance applications, etc. For instance, user traits were predicted using a snapshot of installed apps [113]. App logs of 200 participants were collected, and the applications were grouped based on app purpose. An SVM classifier was employed to successfully infer an individual's religion, marital status, whether the user is a parent of small children, and their mother tongue. Personality traits were identified using individuals' app adoption. Two thousand forty-three Android users app installation logs were used to identify Big Five traits based on the categories of apps available on the

Google play store [125]. Big Five personality values were binned into three classes low, medium, and high and modeled the personality using a random forests classifier. The model was used to predict the personality traits with 50% success rate; however, they only considered app adoption, not usage.

Phone call behaviors have also been used to estimate the Big Five personality traits of an individual [31]. Using the phone logs of 39 participants, call/SMS related features like call duration, time of calls, and the number of texts messages sent/received were extracted. These extracted features were used to construct the social communication network. Supervised learning method based on SVM was used to predict the Big Five traits with mean squared errors ranging from 0.73 to 0.86 on a 7 point scale. Along with the basic mobile phone information available to all mobile phone service providers, GPS data was also used to predict peoples' personality [30]. Standard carrier logs such as phone calls, text messages sent and received were collected from 69 participants. Using this information, the entropy of calls and texts and also inter-event time between texts/calls was calculated. GPS data was used to find the radius of gyration, distance traveled each day, and the number of places visited. Self-reported Big Five personality traits were binned into three classes: low, average, high and an SVM classifier was employed. Using ten-fold cross-validation, they could identify Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism traits with accuracies 49%, 51%, 61%, 51%, 63% respectively. They reported that Extraversion and Agreeableness traits correlate with the entropy of participants' contacts. The variance of the times between phone calls correlated with Conscientiousness trait. A similar study was conducted with large sample size but also using Bluetooth sensor data [90]. They collected mobile data such as telecommunication data (calls and texts), GPS data, and Bluetooth sensor logs from 636 students for 24 months to identify Big Five personality traits. Using this data the following features were extracted: face to face contacts or physical proximity to others using Bluetooth signal strength, geo-spatial mobility patterns and text messages, calls, and social network friends' contacts lists. Big Five trait values were converted into three classes low, medium, and high. SVM was used to model the personality, and it successfully identified only Extraversion.

Touch screen swipe behaviors have been used to identify the personality [2]. Based on the touch screen swipes collected from 98 participants they extracted touch/swipe related features such as average velocity, mean pressure and mean finger area, and other features. Using the self-reported Eysenck personality questionnaire and extracted features they employed KNN and random forests classifiers to predict the Extraversion and Neuroticism traits with an average accuracy of 62.9%. Smartphone data such as call logs and app logs and Bluetooth scans were used to predict the Big Five Personality traits [24]. This is one of the closest studies to this work. They collected call logs, SMS logs, Bluetooth scans, and application usage from 83 participants for 8 months. Using this data they extracted various features including the number of different Bluetooth ids seen, the average duration of incoming and outgoing calls, number of unique contacts communicated with, average SMS length, number of SMS sent and received, and usage of various applications like Internet, calendar, and office. Personality traits were binned in two classes low and high based on the median value in each trait and an SVM binary classifier was trained to predict Big-Five personality traits

Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism with accuracy 69.3%, 74.4%, 75.9%, 69.6%, 71.5% respectively. They reported that Internet correlates with Extraversion, Conscientious people are less likely to use media applications, and Extroverts spend more time on phone calls.

Smartphone sensor data was also used in assessing emotional stability, happiness, and curing mental disorders [14, 19, 52, 78, 110, 121]. Students' happiness was estimated using smartphone sensor data [63]. Smartphone data such as location data, call, SMS patterns, and physiological data including electrodermal and physical activity using Affectiva Q-Sensor were collected from 68 students for one month. Sensor data was complemented with daily surveys about students' behaviors, activities, and well-being. They modeled the problem as a binary classification problem with two classes *sad*, and *happy* and applied Gaussian mixture models and ensemble methods to predict happiness with 70.2% accuracy. They reported that time spent outdoors correlates with the happiness. Moodscope, an application for sensing the mood based on smartphone usage, was built and successfully demonstrated that mood can be inferred from sensor data [78]. Using application usage logs, phone calls, SMS, email messages, and self-reported mood states (4 times a week) from 32 participants and a multi-linear regression model they could successfully infer participants' mood with an accuracy of 66%. Researchers conducted a large study with 18000 participants and demonstrated that users' mood could be predicted using smartphone features related to location, activity, and sociability with an accuracy of 70% [114]. Academic performance was found correlated to automatic sensing data, and mental well-being [121].

3.6 Rationale for Our Study

There are several limitations to the previous studies.

- Several studies relied on questionnaires/surveys to collect the data, providing psychological insight potentially inaccessible directly from behavior [1, 9, 37, 55, 58, 75, 83, 97, 107, 108].

- Some implementations required data (the number of phone calls initiated or received, call response rate, phone contacts, SMS usage details ranging from number of messages sent to message response rate and social networking activities) that are proprietary and only available to service providers and social networking companies [1, 8, 9, 24, 30, 31, 45, 45, 102, 108, 115].

- Some approaches require analyzing personal email and social network activities (for example Facebook likes, number of friends), which might contravene privacy laws or are contrary to research ethics guidelines [8, 31, 45, 45, 102].

- Most of the contemporary studies modelled human personality using statistical analysis and classification models by converting the standard continuous personality trait values into discrete bins [2, 9, 24, 30, 63, 83, 90, 101, 107, 108, 115, 124, 125].

- The majority of the studies that used machine learning models to predict the human personality lack proper presentation of the model performance and simply expressed/claimed results in terms of statistical metrics, without providing baseline models for comparison [24, 30, 90, 101, 124].

In contrast to these, in this study,

- An opportunistic and unobtrusive sensing method was employed to collect the data and focused on more readily accessible smartphone data, such as app use (but not internal app data) and SMS engagement (but not content). Only the duration of app usage was collected (for example time spent in using messaging apps), which is readily available to any app developer with the users' consent, removing any dependence on social networking companies or network service providers.

- WiFi sensor data was used to establish the location and represent movement, providing a location-based feature absent from previous work. Additionally, this is one of the first attempts to make use of location-anchored WiFi traces, as recorded by smartphones, to identify behavioral traits of individuals across multiple validated personality dimensions, not just general happiness and attitude. In particular, location was used to study only those periods of the day during which participants were in a common space, a university campus. This allowed having some control over the comparative aspects of their movement, schedules, and interaction (with space and each other) unconstrained from time.

- Most of the contemporary studies modeled the prediction of personality traits as classification problems by converting the continuous personality trait values into classes based on value ranges, whereas in this study, regression models were used, thus more accurately representing personality in the Big Five model.

- Finally, results are presented with a more fulsome analysis of model performance, including a suitable baseline model, and analysis of fit behavior through residual analysis.

Chapter 4

Experimental Setup - Data Collection

4.1 Study Design and Data Collection

4.1.1 Participants

Participants were drawn from the Social Sciences research study pool at the University of Saskatchewan, Canada. Data collection was piggy-backed on three studies to facilitate a larger sample size. The first study was run from 25-09-2016 to 25-10-2016 consisted of 84 participants. The second study ran from 30-10-2016 to 09-12-2016 and consisted of 91 participants, and the third study ran from 06-02-2017 to 07-03-2017 with 200 participants. The total number of participants from all three studies were 375 of which 103 identified as male, 165 as female and the remaining participants didn't disclose. 127 participants were aged 17-24 years, 123 aged 25-34, and 19 participants were over 35.
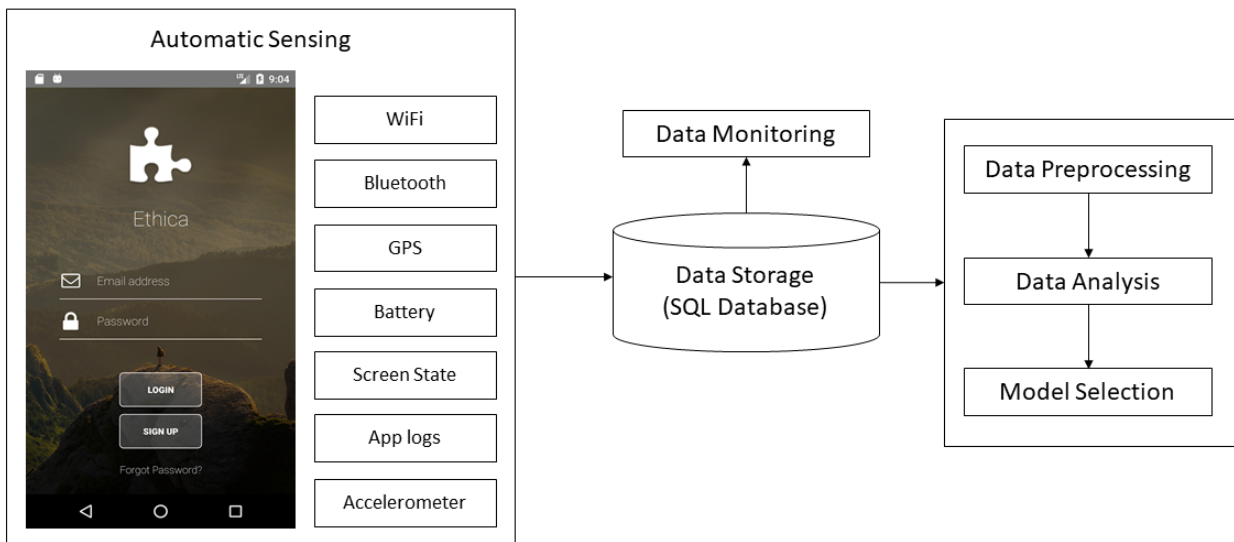

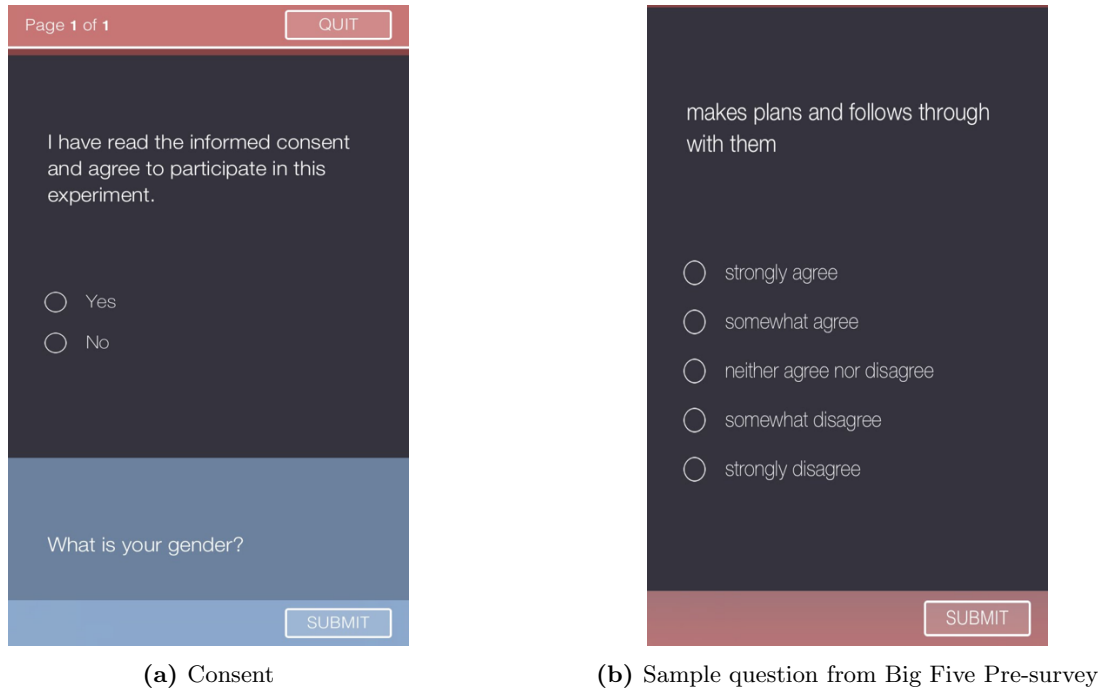**Figure 4.1:** Graphical Representation of Study Setup

(a) Consent

(b) Sample question from Big Five Pre-survey

Figure 4.2: Study Consent and Pre-survey

4.1.2 Study Setup

Data was collected using Ethica Health, an automatic data collection software application from Ethica Data Systems that runs on the Android and iOS smartphone operating systems [38]. Ethica Health collects data from smartphone sensors, including data about location, movement, communication and app usage, social proximity, and momentary ecological assessments (pop-up surveys). Figure 4.1 shows the study design. Ethica Health was installed on participants' phones and set to collect data for one minute every 5 minutes from several sensors, including WiFi, Bluetooth, battery logs, screen state (on/off), accelerometer readings, GPS locations, and which apps were used by participants and for how long. Before data collection started, participants provided informed consent according to university ethics approval, answered the Big Five questionnaire [65] on their phones, and completed a short demographic survey as shown in Figure 4.2a and Figure 4.2b. Participants may have answered other questions not related to this study as part of the other studies this data collection was associated with, but those datasets were not included in this analysis.

Participants were instructed to enable the GPS, WiFi, and Bluetooth sensors at all times during the study. To provide some modicum of privacy, Ethica Health also provides an option (the snooze button) to turn off the data collection for a period of one hour at a time. If any of the sensors are accidentally turned off, the Ethica Health application automatically notifies the participants as shown in Figure 4.3. Here Ethica Health notifies the participant that Bluetooth and GPS sensors are turned off.

Ethica Data Systems (EDS) uses opportunistic sensing method for the data collection and uploads the collected data to the server based on user settings for preferred network usage to data upload. EDS collects
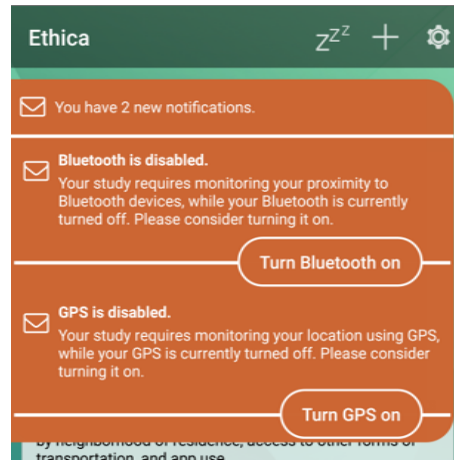
Figure 4.3: Ethica Health Notifications

Figure 4.4: EDS Web Dashboard

data on a duty cycle basis for one minute every five minutes. If the phone is functional, and the EDS app is running, battery level records are guaranteed to be recorded. Data quality can therefore be estimated by counting the number of duty cycles for which battery data is available. During the study, ongoing data collection can be monitored using the EDS web dashboard as shown in Figure 4.4. In the example provided for the participant with id 513, GPS, accelerometer, and Bluetooth sensor data were not collected adequately. This might be due to accidentally turning off the sensors. In such cases, participants were contacted through email via the study pool administrator to check their settings. Collected data was stored in an SQL database. During data preprocessing, participants were filtered based on the proportion of total possible records returned. Participants with fewer than 50% of the total possible records for each study were removed, reducing the 374 participants to 194 participants. Data was preprocessed further and analyzed to select the relevant models for generalizing the data, full details are mentioned in the following sections. Table 4.1 provides the information about various studies conducted and the data collection statistics of the corresponding study.

22

Study	Time line	Partcipants	≥ 50%	≥ 75%	WiFi	App log	Battery	Screen
1	25-09-2016 to 25-10-2016	84	72	67	7,684,192	265,780	13,516	343,979
2	30-10-2016 to 09-12-2016	91	73	55	10,265,197	225,173	17,443	867,629
3	06-02-2017 to 07-03-2017	200	49	45	5,190,487	120,082	11,783	222,513

Table 4.1: Data Collection Statistics and Number of Records collected from various Sensors. % indicates data availability.

4.1.3 Privacy Considerations

Participants can disable data collection by pressing a snooze button ('ZZZ'), shown in Figure 4.3 within the app. Data was encrypted on the phone and transmitted over an encrypted channel. Participants' privacy was maintained by anonymizing each participant's identity with a random user id. The user id map was separated from all other project data so that the data could not be traced back to individuals. Data did not include any internal phone data such as contacts or call logs, SMS logs, or internal app data (email content, Facebook profile data). Wherever possible, data was treated in a manner consistent with the guidelines recommended in [41].

4.2 Data Modeling Steps

Figure 4.5 shows the various steps involved in data preprocessing, analyzing, and modeling the data.

- Initially raw sensor data was collected using EDS software.

- A data compliance filter of greater than or equal to 50% of total possible records was applied to consider participants with the adequate data for the modeling.

- As the study focused on extracting the daily routines for each participant, during stratification the data was grouped based on the user and day. For example, to extract participants' daily app usage-related features using the app log sensor data, first, the data was grouped based on user and date, and then this subset of data was processed to find the app usage of the corresponding participant on a selected day. The same process was applied to other sensors.

- After extracting the features from different sensors for each user on a daily basis, all the individual features were grouped in the aggregate step. For example, first daily app usage features were extracted from app log sensor and campus arrival time was extracted from the WiFi sensor separately, then the extracted features were grouped since both features belong to the same participant.

- After extracting and aggregating features from the sensors, the preprocessed data was analyzed to find correlations between the dependent variables and independent variables. Based on the results from the data analysis step an appropriate machine learning model was selected.
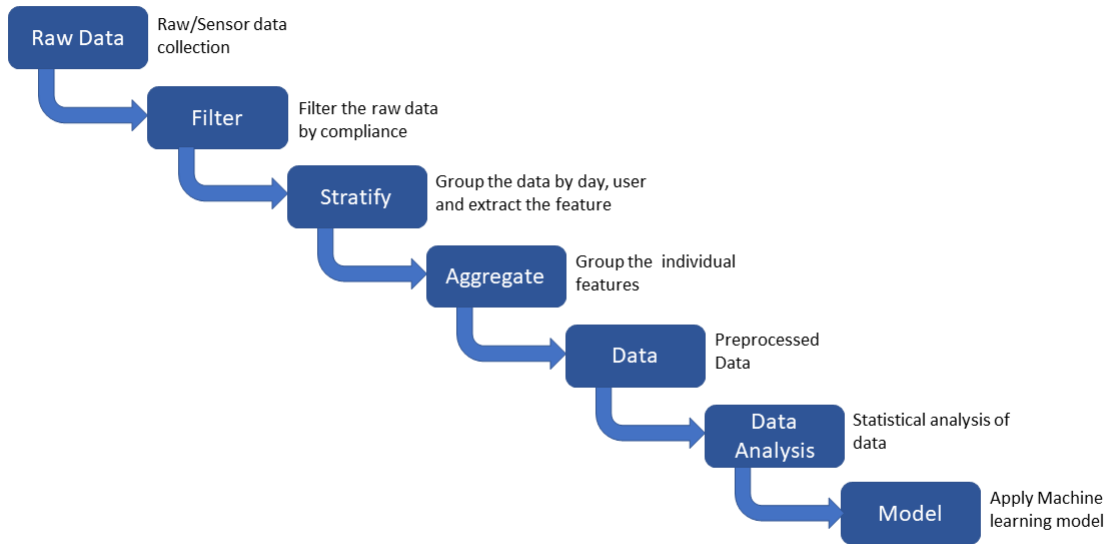
Figure 4.5: Data Pre-processing and Modeling Steps

4.3   Feature Extraction

From the available sensors, the analysis was based on WiFi (as a low-power proxy for location), Bluetooth, app usage, screen state data, and battery sensor data. Each of these data streams provided multiple features for modeling. The research was primarily focused on participants' daily activities and spatial behaviors and how they might correlate with personality traits. The unit of observation for the extracted features is person-day.

4.3.1   Big Five Pre-survey

Each participant's Big Five personality traits were calculated from a pre-study survey completed on the participants' phones. In the Big Five Questionnaire [65] participants answer 44 questions about themselves on a scale of 1-5 (*1-strongly disagree*, *2-somewhat disagree*, *3-neither agree nor disagree*, *4-somewhat agree*, *5-strongly agree*): for example *is talkative?*, *is inventive?*. For each Big Five trait, there are 7-9 questions to be answered by participants. In the pre-survey two questions were removed out of 44 questions for simplicity. Removed questions were: *has an active imagination?*, *remains calm in tense situations?*. Below are the pre-survey questions answered by participants in the study. Using the pre-survey responses, personality traits were calculated based on the scoring instructions from [65]. Personality trait values were normalized between 0 and 1 prior to training the machine learning models. The Big Five questionnaire was conducted only once at the beginning of the study.

1. Is talkative?
2. Tends to find fault with others?
3. Does a thorough job?
4. Is depressed, blue?
5. Is original, comes up with new ideas?
6. Is reserved?
7. Is helpful and unselfish with others?
8. Can be somewhat careless?
9. Is relaxed, handles stress well?
10. Is curious about many different things?
11. Is full of energy?
12. Starts quarrels with others?
13. Is a reliable worker?
14. Can be tense?
15. Is ingenious, a deep thinker?
16. Generates a lot of enthusiasm?
17. Has a forgiving nature?
18. Tends to be disorganized?
19. Worries a lot?
20. Tends to be quiet?
21. Is generally trusting?
22. Tends to be lazy?
23. Is emotionally stable, not easily upset?
24. Is inventive?
25. Has an assertive personality?
26. Can be cold and aloof?
27. Perseveres until the task is finished?
28. Can be moody?
29. Values artistic, aesthetic experiences?
30. Is sometimes shy, inhibited?
31. Is considerate and kind to almost everyone?
32. Does things efficiently?
33. Prefers work that is routine?
34. Is outgoing, sociable?
35. Is sometimes rude to others?
36. Makes plans and follows through with them?
37. Gets nervous easily?
38. Likes to reflect, play with ideas?
39. Has few artistic interests?
40. Likes to cooperate with others?
41. Is easily distracted?
42. Is sophisticated in art, music, or literature?

4.3.2   WiFi

On all Android phones, data for all in-range WiFi routers was recorded, not just the router to which a handshake connection is established. As the data collection was piggybacked on experiments investigating other spatial phenomenon additional data such as GPS location was available but not employed to minimize the intrusion into participants' privacy. Over 23 million (23,139,876) WiFi sensor records were collected from all participants. Table 4.3 shows a sample of five WiFi records. Each WiFi record contains user id, device id, bssid: the router mac address, ssid:network name, record time, and RSSI level value which indicates the proximity of smartphone to the router. To provide Internet service on campus, the University of Saskatchewan has installed more than 1200 WiFi routers. The geographic coordinates of each router including its room number, and building name were available, as shown in Table 4.4.

Because all participants were undergraduate and graduate students from the same university, there was a known spatial anchor point (the university campus) for all participants. Using the known MAC addresses and geographic coordinates for each campus WiFi router, each participant's campus arrival time, campus departure time, time spent on campus, and which buildings were visited on campus were established on a

user id	device id	record time	bssid	level	ssid
551	d98f3ee01cff8e0e	2017-02-12 04:34:38.799000	b8:16:19:e0:54:9d	-69	SHAW-EF0B88
777	112dde4393bc99b8	2017-03-01 13:46:27.122000	28:93:fe:c9:14:ae	-59	uofs-secure
1038	b9f53bef54085973	2017-02-21 23:15:04.474000	02:fc:8d:ba:96:70	-54	SHAW-BA9670-5G
1049	1116373d4fadd09b	2017-02-22 11:13:09.768000	34:bd:c8:7e:b9:64	-65	uofs-public
1302	26b1b610bb1621c6	2017-02-24 23:22:20.510000	10:9f:a9:6d:68:b3	-80	SASKTEL0042

Table 4.3: WiFi Sensor Data

Building	MAC	Location	Floor	Longitude	Latitude
Arts	34:bd:c8:ac:ce:90	Rm 251	2	-106.6376966	52.13231529
Athabasca Hall	00:21:a0:36:69:f0	Rm 170	1	-106.6334167	52.12982611
Place Riel	40:f4:ec:b3:3e:60	Rm 220	2	-106.6359533	52.13081822
Thorvaldson	04:da:d2:a7:45:00	Rm 110	1	-106.6361538	52.13251558
Saskatchewan Hall	00:1f:ca:50:9e:a0	Rm 21	0	-106.6345734	52.13002258

Table 4.4: WiFi Routers Location Information

daily basis. The number of unique WiFi routers visited around the city for each day in the study were also calculated. Routers installed nearer to university transit terminal (Place Riel) and university residences were not considered for campus related features as it was possible that participants living close to or at the university might have take the bus to commute other places.

WiFi records with RSSI values greater than -60 dBm were considered i.e. indicating (for us, colloquially established) proximity of approximately 30 meters or less to the router. Campus arrival Time was estimated as the earliest time in a day where a participant's phone recorded a router with a university SSID; the last time a campus router is seen was used as departure; calculating the difference gives the time spent on campus. The room number and building name associated with each router were used to establish the range of buildings and places visited on campus and different WiFi routers seen around the city. University arrival/departure times were selected because it was hypothesized that the regularity of arrival/departure times would correlate with Conscientiousness. It was anticipated that time spent in and different buildings visited on campus would correlate with Conscientiousness as people who score high on the Conscientiousness trait tend to plan their daily activities. Openness would correlate with different WiFi routers seen in the city, implying that participants visit new places.

$$\text{Campus entry time } = \text{first(campus routers times)}$$
$$\text{Campus departure time } = \text{last(campus routers times)}$$
$$\text{Time in Campus } = \text{Campus entry time } - \text{Campus departure time}$$

(4.1)

4.3.3 App usage logs

For application (app) usage data, 611,035 records were collected. Table 4.5 shows sample records. Collected information included user id, device id, record time, app name, and fg_time_ms: foreground/active app usage time (milliseconds) of the application. Using this data, the total duration of an application use each day, the cumulative usage for each application, was calculated. Every application has a specific purpose, but certain sets of applications can be grouped together based on common functionality. For example, Viber, Whatsapp, and Skype can be grouped into messaging applications. Likewise, there are many variants of similar applications used for music, email or gaming. To compare and generalize the app usage information, applications were grouped as shown in Table 4.6. Applications were manually categorized into groups. For example applications with names 'com.android.phone', 'com.sec.phone', 'viber', 'messaging', 'whatsapp' were categorized into Social Media/call/SMS apps. Similarly applications with names 'camera', 'google maps', and 'chrome' were categorized into Camera/Maps/Internet apps. Camera, Maps and Internet related applications were grouped because it was expected that Camera, Internet, and Map usage would correlate with the Openness trait. For each category, the app usage of that category was calculated as a percentage of total app usage, as shown in Equation 4.2. Social Media, Calls, and SMS usage have been shown to correlate with Extroversion [9, 25, 83, 108]. Scheduling, Office, Music, and Weather app usage have been shown to correlate with Conscientiousness [24].

$$\text{Social Media, Calls, SMS usage} = \frac{\text{Social Media, Calls, SMS usage}}{\text{Total App Usage}} \tag{4.2}$$

user id	device id	record time	app name	fg_time_ms
943	a63c280bfb7fee3f	2017-02-06 01:55:01.293000	com.whatsapp	12613945
1044	23a8f7ecc0a54c1	2017-02-13 07:52:19.845000	com.sec.android.app.camera	69640
1078	b941129ca5c8d842	2017-02-06 12:47:42.090000	com.google.android.apps.maps	2710
1301	b8cedcc9bf3b8932	2017-02-07 00:01:09.308000	com.google.android.calendar	11782
1301	b8cedcc9bf3b8932	2017-02-07 00:01:09.318000	com.spotify.music	144867

Table 4.5: App logs

Category	App Names	Correlated Trait
Social Media, Calls, SMS	calls, SMS, whatsapp, viber, etc.	Extraversion
Camera, Maps, Internet	camera apps, navigation apps, browser, etc.	Openness
Scheduling, Office Tools, Weather	organizer, calender, e-mail, dictionary, weather apps, etc.	Conscientiousness
Media Apps	music, videos player, photos viewer, etc.	Conscientiousness

Table 4.6: App Category and Correlated Trait

4.3.4 Battery Sensor

From the battery sensor, 42,742 records were collected. Sample battery sensor data is shown in Table 4.7. Data contained information about user id, device id and plugged state (0- unplugged, 2- charging via USB and 1- plugged to AC). Data also includes several event details: when was the mobile plugged for charging, when was unplugged from charging, start, and end battery levels. For example in Table 4.7 participant with user id 1364 charged the mobile via USB from 2017-02-28 19:26:06 to 2017-03-01 05:08:59. The starting battery level was 45, and end level was 100.

user id	device id	start_time	end_level	end_time	plugged	start_level
551	d98f3ee01cff8e0e	2017-02-07 18:03:43	60	2017-02-07 18:21:50	1	47
551	d98f3ee01cff8e0e	2017-02-07 17:20:08	37	2017-02-07 17:51:32	1	10
956	43151955553b67c	2017-03-04 05:56:33	100	2017-03-04 11:01:08	1	58
1299	2ffcbac9a8426795	2017-02-10 07:33:07	89	2017-02-10 09:56:01	0	100
1364	80aa56fba5dcf8dc	2017-02-28 19:26:06	100	2017-03-01 05:08:59	2	45

Table 4.7: Battery Sensor Data

From this data, the extracted features were daily battery charging duration and if and when charging was occurred. Sometimes it is possible that participants may charge their mobile multiple times in a given day, for example, participant 551 charged their mobile twice on Feb 2. In such cases, the most frequent time over the course of the study was considered as the most representative of the multiple events as shown in Algorithm 1. It was hypothesized that the regularity of battery charging times would correlate with conscientiousness because conscientious people like to do things in an organized manner [48].

Result: daily charge time

Input : daily charge times, study charge times;

if *count(daily charge times) > 1* **then**

    for *time ←daily charge times* **do**

      | probability of time in study charge times

    end

    daily charge time = daily charge time with highest probability;

else

    daily charge time = daily charge times[0] ;

end

Algorithm 1: Algorithm to find the daily charging time

4.3.5 Bluetooth

From the Bluetooth Sensor, 2,247,499 records were collected. Sample data is shown in Table 4.8. User id and device id were helpful to identify the participants' device. All the nearby Bluetooth devices were logged by the sensor data not just the handshake connection established devices. Devices include printers, laptops, and mobiles, etc. 'dev class' column contains the information about the type of recorded device. For this study, only mobile device classes were considered for analysis. Mobile device classes have the hexadecimal id of any in: [01020c,50020c,52020c,58020c,5a020c,62020c,70020c,72020c,78020c,7a020c].

Using the Bluetooth sensor data the number of contacts for a smartphone with smartphones was counted i.e., the number of contacts of each individual participant was proximate to other individuals. It was hypothesized that the number of contacts correlates with the Agreeableness trait. The number of data points extracted from the Bluetooth sensor was insufficient compared to other sensor-related features, this reduced the total number of data points after aggregation with other sensor features. Therefore, further analysis of Bluetooth features was not considered in this study.

user id	device id	record time	dev class	mac	rssi
551	d98f3ee01cff8e0e	2017-02-12 04:34:38.799000	1f00	b8:16:19:e0:54:9d	-69
777	112dde4393bc99b8	2017-03-01 13:46:27.122000	5a020c	28:93:fe:c9:14:ae	-59
1038	b9f53bef54085973	2017-02-21 23:15:04.474000	380104	02:fc:8d:ba:96:70	-54
1049	1116373d4fadd09b	2017-02-22 11:13:09.768000	0x5a020c	34:bd:c8:7e:b9:64	-65
1302	26b1b610bb1621c6	2017-02-24 23:22:20.510000	0x72020c	10:9f:a9:6d:68:b3	-80

Table 4.8: Bluetooth Sensor Data

4.3.6 Screen State

From screen state 1,434,121 records were collected, sample data is shown in Table 4.9. Screen state data contains information about the screen on/off state along with a time stamp. Using this data each participant's daily(active) smartphone usage was extracted. Continuous screen ON and OFF states were calculated for a single day and summed which gave the participant's active mobile duration for a day. It was hypothesized that Screen State negatively correlates with Conscientiousness as conscientious students are studious and less likely to use mobile for longer periods of time and that Screen state positively correlates with Extroversion. As of our knowledge, this is the first time the Screen state sensor is used in estimating human personality traits.

$$\text{Active Smartphone usage} = \frac{\text{screen ON}}{\text{screen ON} + \text{screen OFF}} * 100 \%$$

user id	device id	record_time	state
1334	9dc1295da64818e5	2017-02-08 08:29:22.000000	1
1334	9dc1295da64818e5	2017-02-08 08:29:42.000000	0
1346	329155801335c999	2017-02-18 23:48:18.000000	0
1346	329155801335c999	2017-02-18 23:49:32.000000	1
514,	19b16d601cf97b8f	2017-02-07 17:25:41.000000	1
514	19b16d601cf97b8f	2017-02-07 17:28:28.000000	0

Table 4.9: Screen Sensor Data

Table 4.10 shows a summary of extracted features from different sensors; Table 4.11 shows a sample of the final dataset including dependent and independent variables. Time values are converted into decimal numbers, for example, 8:30 am is converted to 8.5 and 8:30 pm is converted to 20.5. Each participant had a maximum of 43 records; one record daily consisting of spatial features and their personality traits. Some participants might have less than 43 records due to technical issues, battery drain, or consciously turning off or snoozing the EDS app. Overall 1929 data points were extracted from 169 participants. For the study, only participants whose data points spanned more than ten days were considered. After filtering, there were 1461 day records from 80 participants. All our study data is maintained in an SQL database. A machine with Intel Core i7 processor, 16 GB RAM and Python 3.6 is used for feature extractions and running the machine learning experiments.

Data Source	Extracted Feature	Aggregation	Units
WiFi Sensor	Campus entry time	Daily	hh.mm
	Campus leave time	Daily	hh.mm
	Time spent in campus	Daily	hh.mm
	Different buildings visited on campus	Daily	number
	Different WiFi routers seen in city	Daily	number
App usage Data	Social Media, Calls, SMS usage	Daily	% of total usage
	Maps, Internet, Camera usage	Daily	% of total usage
	Scheduling, Office, Weather apps usage	Daily	% of total usage
	Media apps (music,video) usage	Daily	% of total usage
	Other apps usage	Daily	% of total usage
Battery Sensor	Most probable battery charge plugin start time	Daily	hh.mm
	Total Duration of Plugin	Daily	hh.mm
Screen State Sensor	Active smartphone usage	Daily	%
Pre-survey	Big Five Personality Treats (Ground Truth)	Once Per Study	number (0-1)

Table 4.10: Summary of Data Sources and the Extracted Features

ID	Date	MA	OA	CT	CD	ET	LT	TC	W	B	AP	SA	CA	SCA	E	A	C	N	O
506	09-29	0.03	0.37	13	1.17	8.6	17.6	9	125	3	81	0.29	0.02	0.3	0.78	0.97	0.83	0.14	0.75
514	10-11	0	0.32	4	2.85	8.9	14.7	5.8	26	3	88	0.61	0.02	0.07	0.53	0.56	0.69	0.54	0.53
519	10-17	0.03	0.78	7	1.5	8	15.8	7.8	98	6	84	0.16	0.03	0	0.59	0.81	0.69	0.5	0.75
532	10-24	0.04	0.28	1	1.83	13.6	13.7	0.08	8	1	76	0.67	0.02	0.02	0.81	0.81	0.69	0.61	0.72
534	09-26	0.04	0.98	0	6.92	10.7	23.9	13.2	32	3	86	0.07	0.04	0	0.91	0.72	0.83	0.71	0.86

Table 4.11: Sample Data of 5 records

MA- media apps usage; OA- other app usage; CT- charging time; CD- battery charging duration; ET- campus entry time; LT- campus leaving time; TC- time in campus; W- Wifi routers seen in city; B- different buildings visited on campus; AP- active phone use; SA - social chat, calls, and sms apps usage; CA- camera, maps, and internet usage; SCA- scheduling, office tools, and weather apps usage; E- Extraversion; A- Agreeableness; C- Conscientiousness; N- Neuroticism; O- Openness

4.4 Data Analysis

In this section baseline distributions of all independent and dependent variables are provided to help establish the form of the dataset.

4.4.1 Univariate analysis

Independent Variables - Input Features/Extracted Features

Figure 4.6 - Figure 4.8 shows the histogram distributions for all extracted input features and Table 4.12 shows the aggregate feature statistics. For plots Campus Entry Time (Figure 4.6a), Campus Leaving Time (Figure 4.6b) , Battery Charge Start Time (Figure 4.6c), Battery charging duration (Figure 4.6d), and Time Spent in School (Figure 4.6e), the x-axis represents the hours of day and y-axis represents the frequency. As shown, Campus Entry times have peaks around 10 am and Campus leaving time has a mean value at 4.27 pm. Surprisingly, there are a few outliers present for Campus Entry time and Campus Leaving time, these are because some participants carried study phones along with their personal phones and they might have left their phones on campus or at home. Daily battery charge start times peak at 10 pm, and 12 am, which is likely because the majority of participants plugin before going to sleep. Average battery charging duration time is 4.4 hours and has a higher standard deviation of 3.4 hours, as many values are near the lower bound. Average time spent on campus is 5 hours, and many of the participants are in 0-3 bin indicating brief stays on campus.

The x-axis for histograms of Different Building Visited on Campus (Figure 4.6f) and Number of Different WiFi routers (Figure 4.7a) represents Number and y-axis represents frequency. On average the number of buildings visited on campus is 5 and the peaks are evident at 6. In the case of WiFi routers contacts, 190 routers were visited on an average each day, likely because of the density of routers on campus.

For features related to app usage and Smartphone usage, the x-axis represents the percentage of usage and y-axis represents the frequency. Most of the features extracted from app usage logs follow an exponential distribution having heavy tails. Media apps usage (Figure 4.7b) and Scheduling, Office and Weather apps usage (Figure 4.7c) peaks are observed around 0-10%. This is because participants tend to turn on the music apps in the background and Ethica Data System logs only the foreground app usage time. A similar trend is observable for Scheduling, Office tools, and Weather apps usage as individuals open and closes the apps such as calendar and Google weather, and other convenience apps. In the case of Social Media, Call and SMS apps, usage is higher, likely because individuals tend to actively use the apps, ensuring that usage is logged. Active Phone usage has several peaks and follows a multimodal distribution. Average Active Phone Usage is 47.2%, and the standard deviation is 26%.
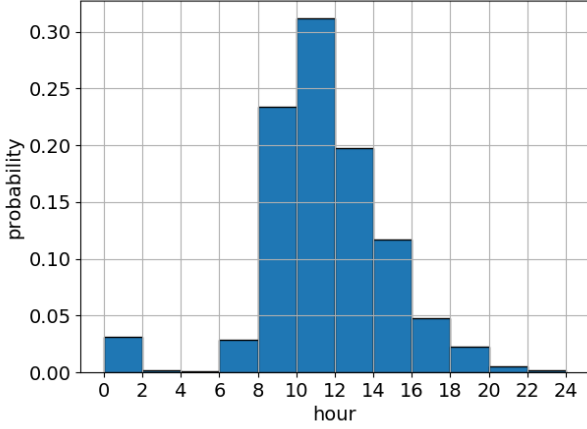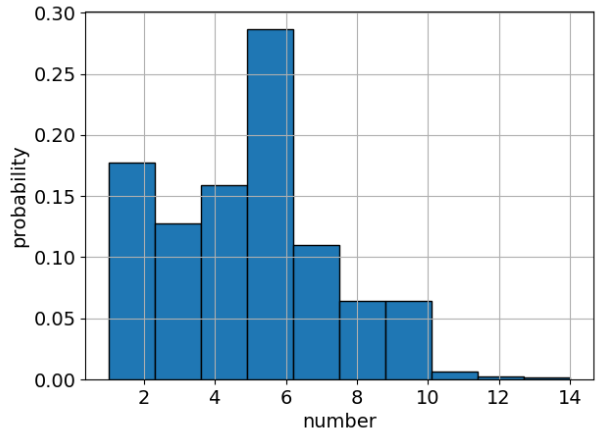
(a) Campus Entry Time
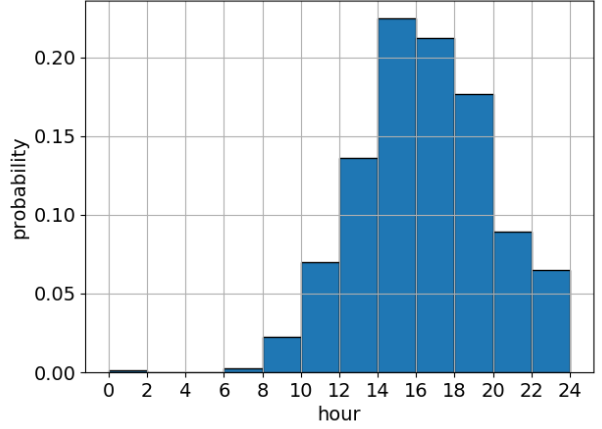
(b) Campus Leaving Time
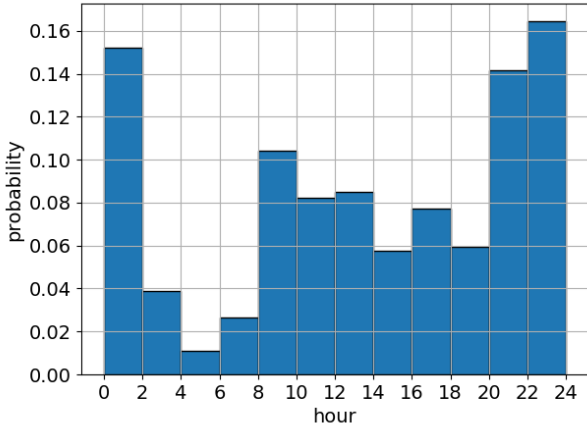
(c) Daily Battery Charge Start Time
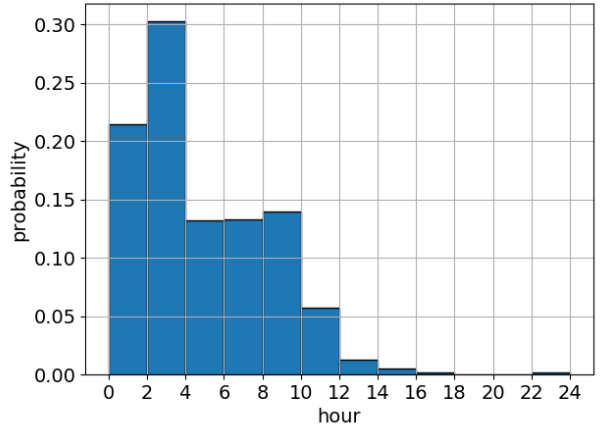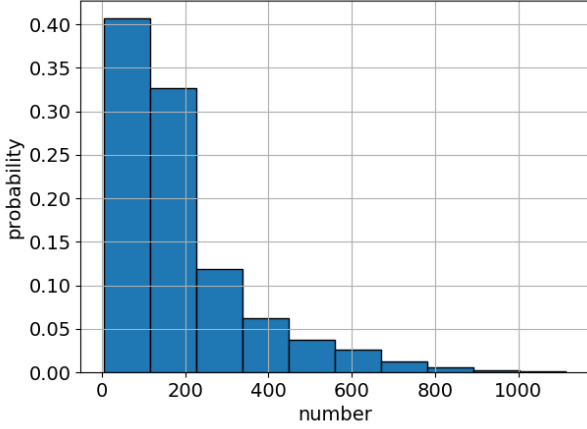
(d) Battery Charge Duration
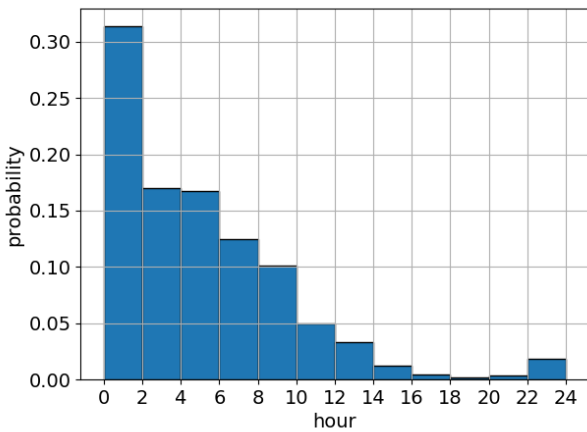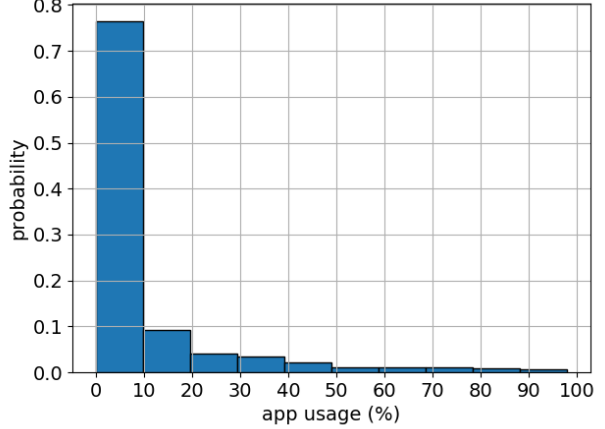
(e) Time Spent in School

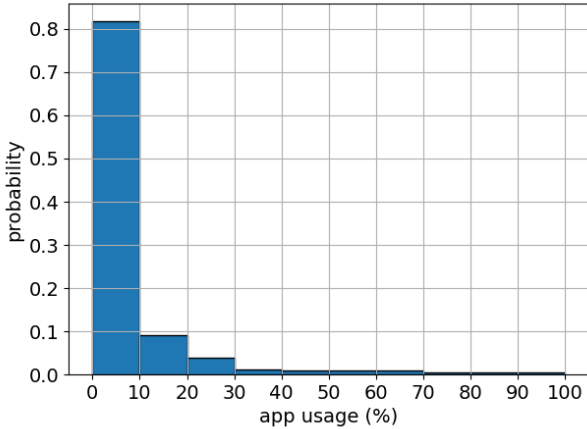**(f)** Different Buildings Visited on Campus

Figure 4.6: Histograms of Extracted Features 1-6
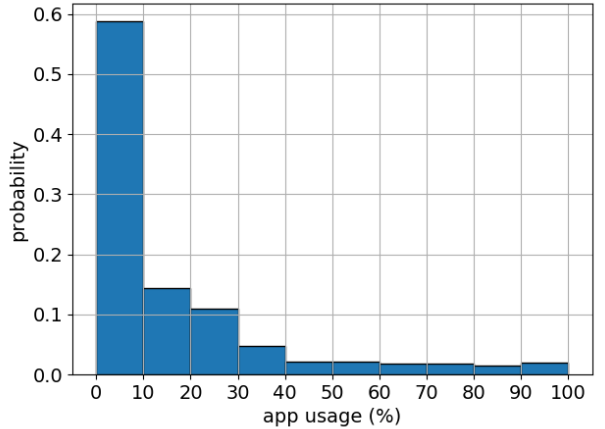
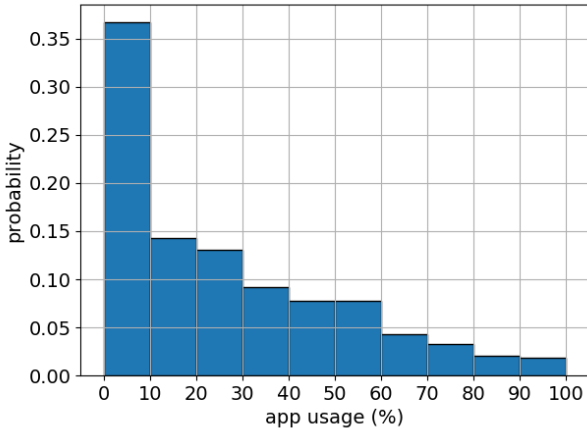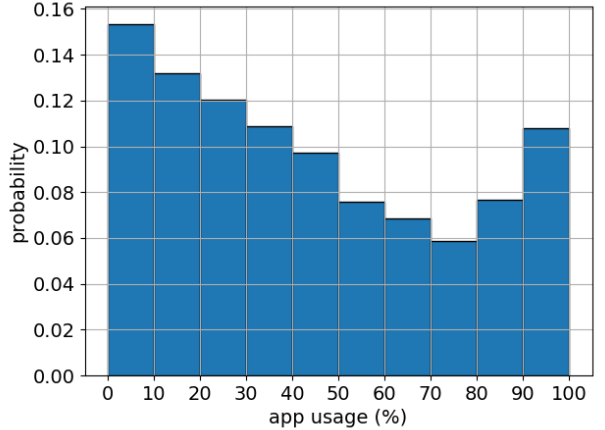**(a)** Different WiFi routers seen around City

(b) Media apps usage

(c) Scheduling, Office and Weather apps usage

(d) Camera, Map and Internet apps usage

(e) SocialMedia, Call and SMS apps usage

(f) Other apps usage

Figure 4.7: Histograms of Extracted Features 7-12
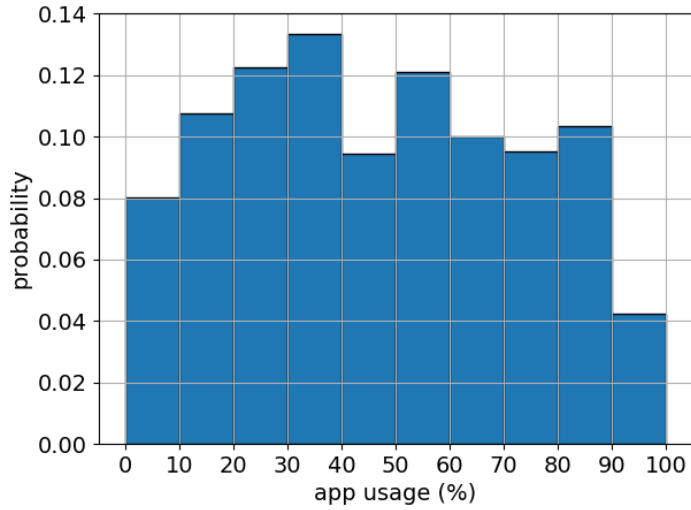
Figure 4.8: Histogram of Active Smartphone Usage

Feature	mean	std	min	25%	50%	75%	max	Skew
Campus entry time	11.4	3.3	0	9.7	11.2	13.3	23	-0.52
Campus leave time	16.45	3.5	0.1	14.2	16.4	18.8	24	-0.025
Time spent in Campus	5	4.7	0	1.2	4.2	7.5	23.9	1.46
Different buildings visited	4.8	2.3	1	3	5	6	14	0.33
Different WiFi routers seen	190.4	159.7	5	85	139	233	1114	1.9
Social Media, Calls, SMS usage	0.25	0.24	0	0.03	0.2	0.42	1	0.91
Maps, Interent, Camera usage	0.15	0.22	0	0	0.06	0.21	1	2.06
Scheduling, Office, Weather usage	0.06	0.13	0	0	0.02	0.07	1	3.7
Media apps usage	0.08	0.17	0	0	0.01	0.1	0.98	2.8
Other apps usage	0.43	0.30	0	0.16	0.38	0.68	1	0.38
Battery charge plugin start time	13	7.6	0	8	14	21	23	-0.36
Total duration of charging	4.4	3.4	0	2	3	7	25	1.11
Active Smartphone usage	47.2 %	26%	0	26%	46.5 %	69%	100%	0.06

Table 4.12: Statistics for Independent variables

Dependent Variables - Big Five Traits

Figure 4.9 and Table 4.13 shows the histogram distributions and statistics for Big Five personality traits. As shown Openness, Conscientiousness, and Agreeableness have higher values which are expected as the traits were collected from university students who are open to new learning, friends and follow schedules for attending classes. Neuroticism has a wide range of values indicating, participants with different emotional stability levels. Overall, the mean values for Extraversion and Neuroticism traits are low compared to other traits because of the more extreme values. Standard Deviations for Neuroticism and Extraversion are high compared to the other three traits as data values are spread out over a wide range of values. Statistics for Openness, Extraversion, and Neuroticism are similar to the previous personality studies conducted on various populations [45, 46]. Conscientiousness and Agreeableness traits have slightly higher values.

Trait	mean	std	min	25%	50%	75%	max	Skew
Openness	0.67	0.1	0.41	0.61	0.66	0.75	0.86	-0.15
Conscientiousness	0.70	0.14	0.39	0.61	0.7	0.8	1	0.3
Extraversion	0.6	0.18	0.22	0.47	0.57	0.72	1	0.22
Agreeableness	0.8	0.13	0.44	0.72	0.81	0.89	1	-0.75
Neuroticism	0.48	0.2	0.03	0.36	0.5	0.61	0.96	-0.19

Table 4.13: Statistics for Dependent variables
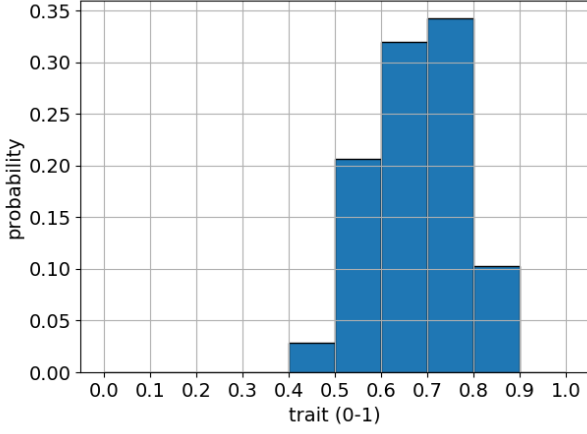
(a) Openness
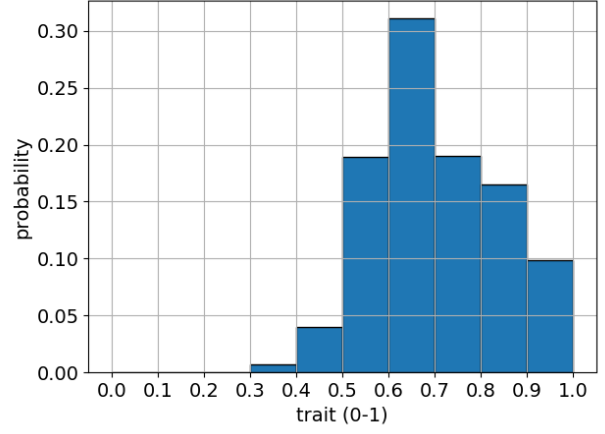
(b) Conscientiousness
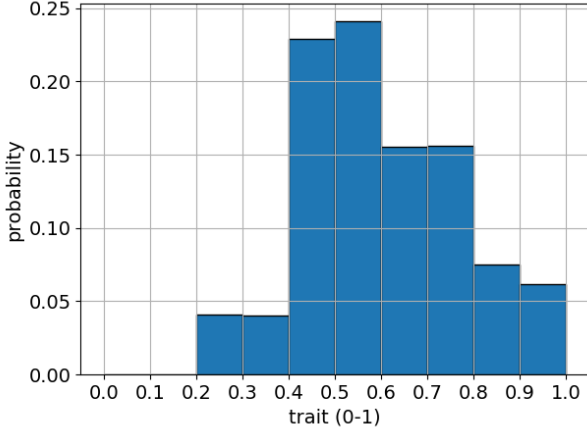
(c) Extraversion
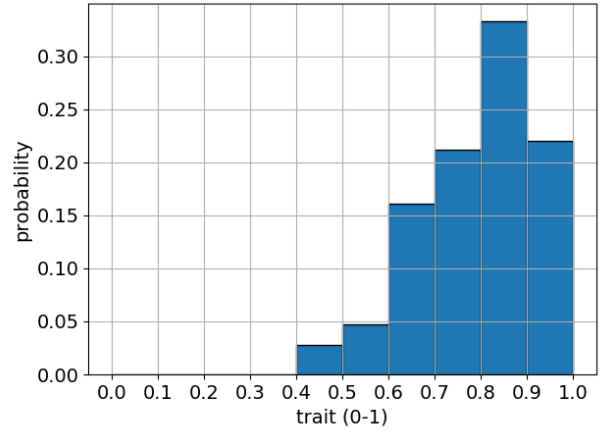
(d) Agreeableness
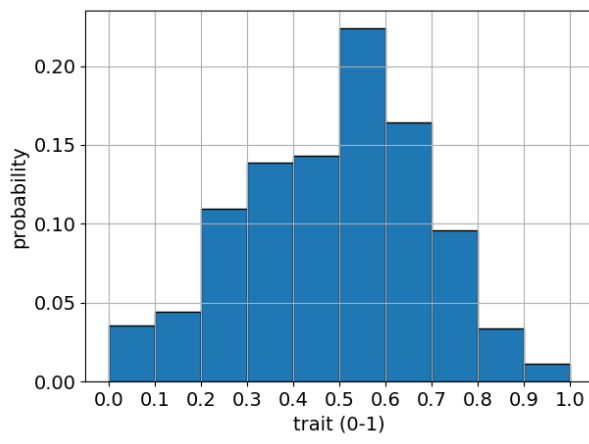
(e) Neuroticism

Figure 4.9: Histograms for Big Five traits

4.4.2 Bivariate analysis

Correlation Between Dependent and Independent Variables

Pearson and Spearman correlations and scatter plot matrices were used to determine if linear or ordinal relationships between measured features and Big Five personality traits exist. Correlation results are shown in Table 4.14 and scatter plot matrices are shown in Figure 4.10, Figure 4.11, and Figure 4.12. Correlations were low ranging from -0.31 to 0.26 indicating that no significant correlations between the extracted features and personality traits, a similar trend was observed in the study [24]. Low correlations were further evident in scatter plot matrices with no significant patterns or trends in the data. This indicates that linear models are unlikely to be sufficient to predict the behavioral traits. Because Pearson and Spearman's correlations provide the bivariate direct correlation between the independent variables (extracted features) and dependent variables (personality traits) but not joint associations of independent variables with the dependent variables, potential models will likely have to be drawn from non-linear associative models. Regression analysis models can capture the joint/conditional associations between collective independent variables and dependent variable. Several experiments were carried out with various machine learning algorithms such as Neural Networks, Decision Trees, Random Forests, Kernel Ridge Regression, and SVMs. Results indicated that neural networks were more able to capture the non-linearity in the data. The ability to fit complex non-linear relationships is a well known property of neural networks [12, 59, 62, 72, 86, 105].

Feature	O		C		E		A		N	
	r	ρ	r	ρ	r	ρ	r	ρ	r	ρ
Campus entry time	0.04	0.06	-0.05	-0.03	-0.01	-0.03	-0.03	0.01	0.04	-0.05
Campus leave time	0.09	0.11	0.12	0.08	-0.08	-0.06	0.16	0.19	-0.14	-0.09
Time spent in Campus	0.04	0.07	0.12	0.08	-0.06	-0.01	0.14	0.14	-0.13	0.01
Different buildings visited	-0.02	0.01	0.09	0.06	-0.20	-0.20	0.02	0.04	-0.02	-0.03
Different WiFi routers seen	0.06	0.11	0.17	0.19	0.01	0.06	0.05	0.05	-0.24	-0.25
Social Media, Calls, SMS usage	0.06	0.12	-0.08	-0.04	0.11	0.14	0.04	-0.01	-0.01	-0.06
Maps, Interent, Camera usage	-0.07	-0.03	-0.10	-0.14	-0.03	0.04	-0.20	-0.31	0.18	0.26
Scheduling, Office, Weather usage	0.11	0.16	-0.14	-0.06	0.11	0.13	-0.25	-0.12	0.21	0.14
Media apps usage	-0.07	0.03	0.02	0.03	0.01	0.08	0.01	-0.03	0.08	0.02
Other apps usage	-0.01	-0.01	0.19	0.14	-0.12	-0.13	0.22	0.21	-0.27	-0.25
Battery charge plugin start time	-0.05	-0.04	0.04	0.04	-0.03	-0.01	-0.05	-0.09	-0.02	-0.04
Total duration of charging	0.02	0.06	-0.05	-0.04	0.02	0.05	-0.02	-0.01	0.02	0.00
Active Smartphone Usage	0.15	0.17	0.04	0.05	0.13	0.13	-0.07	-0.07	0.02	-0.02

Table 4.14: Pearson (r), Spearman (ρ) correlations between extracted features and the Big Five traits

Trait	O	C	E	A	N
Openness	-	0.21	0.43	0.05	-0.088
Conscientiousness	-		0.20	0.44	-0.48
Extraversion	-	-	-	0.13	-0.21
Agreeableness	-	-	-	-	-0.45
Neuroticism	-	-	-	-	-

Table 4.15: Pearson (r) correlations among the Big Five traits
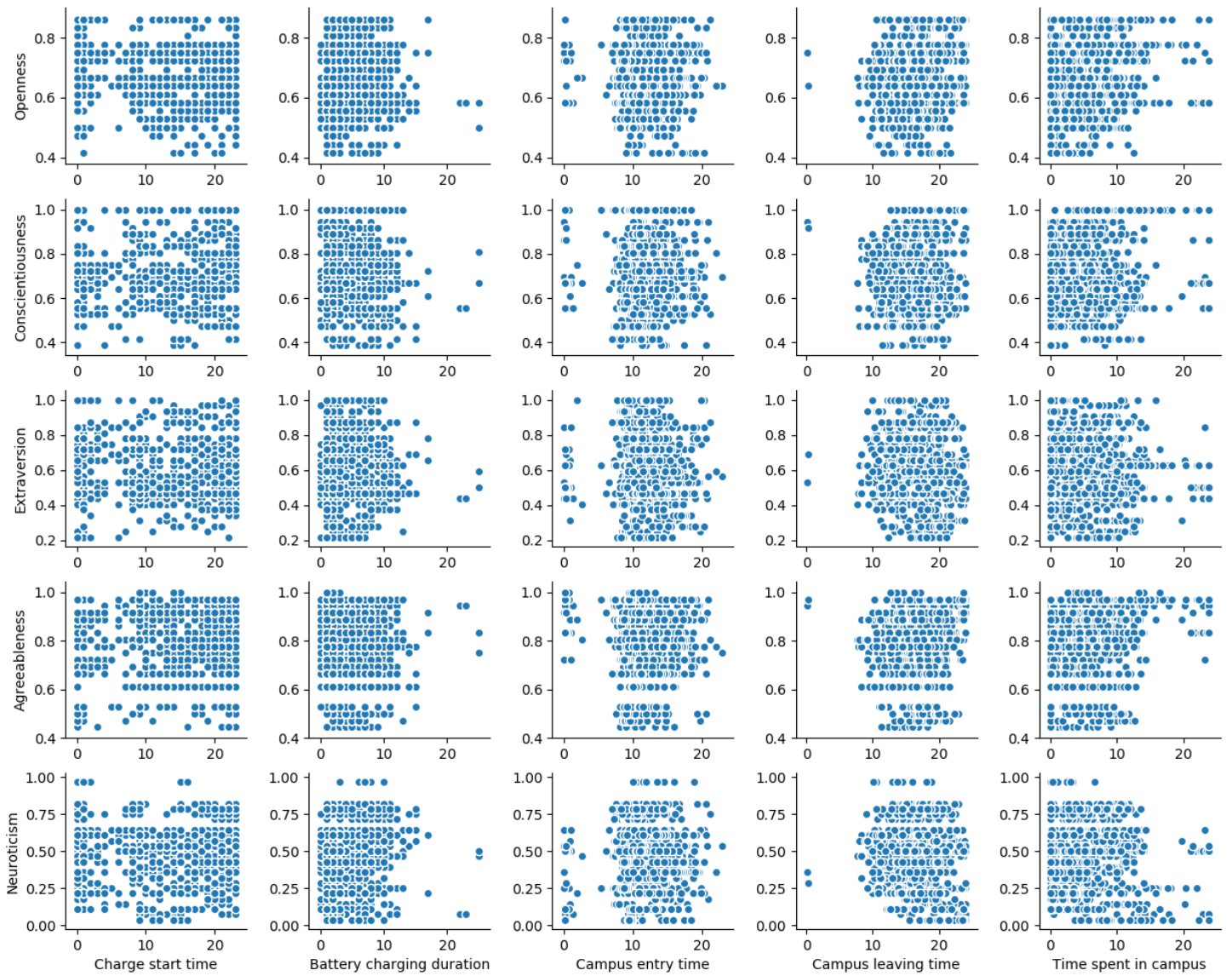
Figure 4.10: Scatter Plot Matrix for the extracted features and the Big Five traits, here x-axis represents the extracted feature and y-axis represents the Personality trait
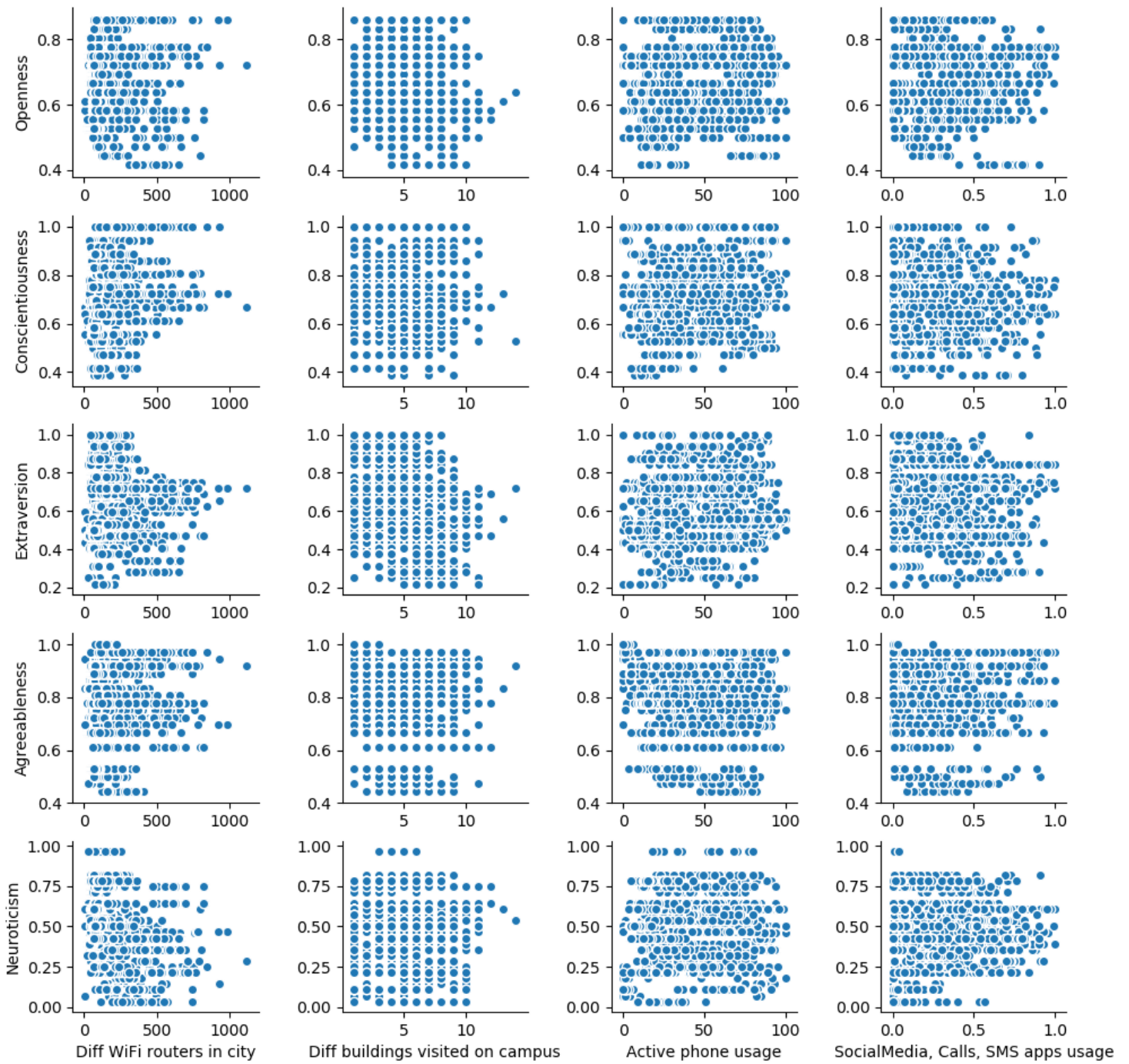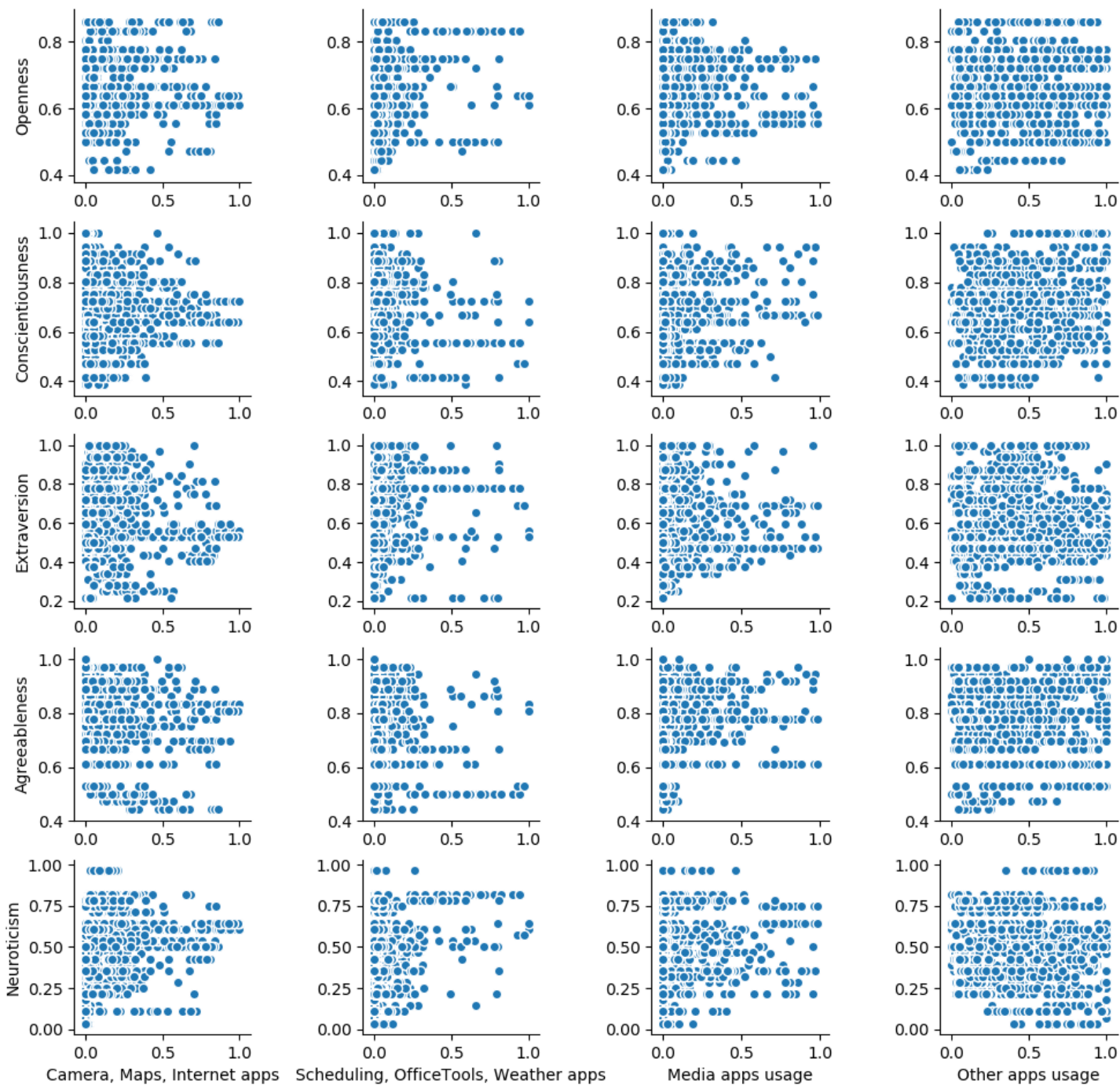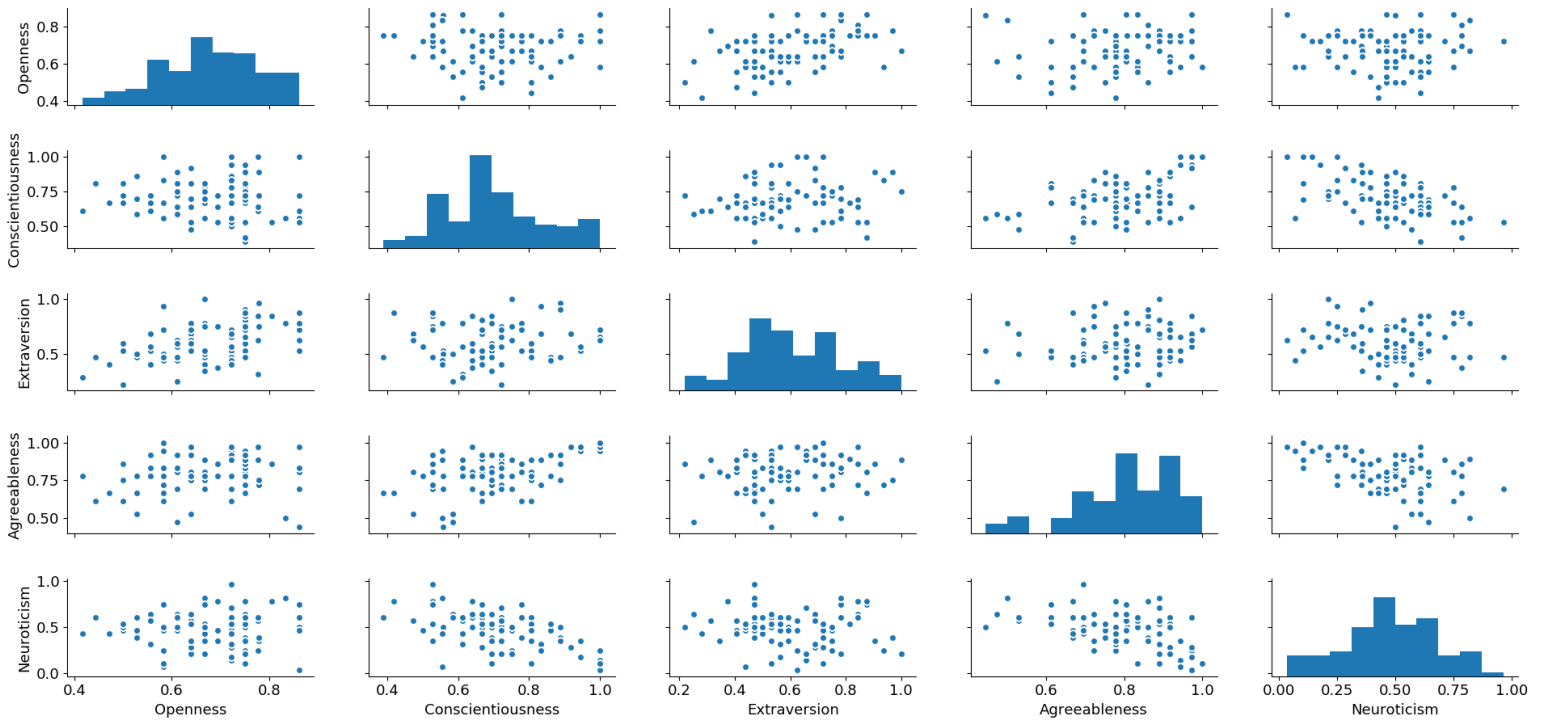
Figure 4.11: Scatter Plot Matrix for the extracted features and the Big Five traits, here x-axis represents the extracted feature and y-axis represents the Personality trait

Figure 4.12: Scatter Plot Matrix for the extracted features and the Big Five traits, here x-axis represents the extracted feature and y-axis represents the Personality trait

Figure 4.13: Scatter Plot Matrix for the Big Five traits, here both x-axis and y-axis represent the Big Five Personality traits

CHAPTER 5

METHODS

Supervised machine learning models were created based on neural networks to predict the Big Five personality traits using the features described in section 4.3. Personality trait values to be predicted were continuous between 0 and 1, making it a regression modeling problem. For the purpose of comparisons and demonstrations of results another regression algorithm, Random Forests, was used. Mean models were used as a baseline model for comparison. Classification models were created based on SVM to establish the guidelines in the presentation of results when using classification machine learning models and to compare to related work. ZeroR models were used as baseline models for comparison.

In all of the experiments, 4-fold cross-validation was used to evaluate the quality of the machine learning models. The entire dataset was divided into 4 folds from which one set was kept aside as a test set, and the model was trained using the remaining 3 sets of data, this method allows for testing against unseen data without potentially biasing the results by randomly choosing a favorable or unfavorable test set. This also makes sure that for every fold 25% of the dataset was kept aside for testing and remaining 75% of the data was used to train the model. Because this is a regression modeling problem, Root Mean Squared Error (RMSE) was used as the metric of success. For classification models, accuracy was used as the success metric. RMSE values and accuracy values were converted into percentage to aid interpretation. The Python programming language was used for creating the machine learning models. Neural networks models were created using the Tensorflow 1.4.0 module, and the Scikit-learn 0.19.1 module was used in the creation of random forests and SVM models.

Detailed explanations about training and tuning the machine learning models are provided in the following sections. Tuning baseline models is not provided, because the baseline models do not have any learning parameters.

5.1 Neural Networks

Preliminary results and analysis indicated that neural networks were able to capture the non-linearity in the data. A classic three layer (input, hidden, output) perceptron or back-propagation network with linear input and output layers was created [44]. Each personality trait was modeled using its own network. A schematic of the neural network model created for the Openness trait is shown in Figure 5.1. The input layer contains 13 neurons representing the extracted input features, and output layer contains one neuron the value of the personality trait, in this case, Openness. Five such models were created for all five personality traits. Neural network models for remaining personality traits are presented in Appendix A.
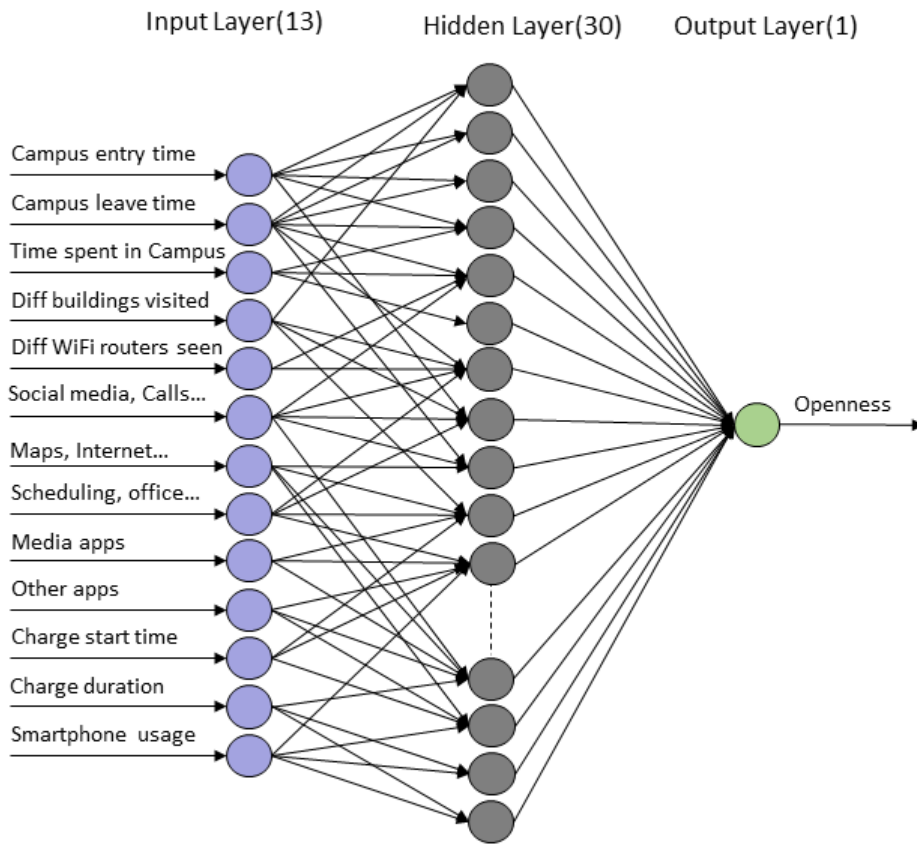
Figure 5.1: Neural Network Model created for Openness

Training the neural network model involved two steps. In the forward pass, all the connections were initialized with random weights and then each hidden neuron values were calculated as shown in Equation 5.1. Hidden node values were squashed using an activation function as in Equation 5.2. The same process was repeated for all hidden nodes. Finally, the output value (personality trait) was calculated based on the weights and hidden node values as in Equation 5.3. Using RMSE, error values were calculated based on the expected output and predicted output as shown in 5.4 and 5.5. In the backward pass, depending on the loss

an optimizer was used to update the weights associated with each neuron as shown in Equation 5.6. The same process was repeated until the predicted output reach acceptably close to expected output.

$$
\begin{aligned}
net_{h1} = \ & w_1 \times \text{Campus Entry Time} + w_2 \times \text{Campus Leave Time} + w_3 \times \text{Time Spent in School} + \\
& w_4 \times \text{Diff Buildings Visited} + w_5 \times \text{Diff WiFi Routers Seen} + w_6 \times \text{Social Media, Calls SMS} + \\
& w_7 \times \text{Maps Internet Camera} + w_8 \times \text{Scheduling Office Weather} + w_9 \times \text{Media apps} + \\
& w_{10} \times \text{Other apps} + w_{11} \times \text{Daily charge time} + w_{12} \times \text{Charge Duration} + w_{13} \times \text{Phone Usage}
\end{aligned}
\tag{5.1}
$$

$$\text{where:} \quad w_1, w_2, w_3 \cdots w_{13} \text{ are weights in the Hidden layer}$$

$$
out_{h1} = Activation\ Function(net_{h1})
\tag{5.2}
$$

$$
\widehat{Openness} = \ w_{11} \times out_{h1} + w_{12} \times out_{h2} + w_{13} \times out_{h3} + \cdots + w_{1n} * out_{hn}
\tag{5.3}
$$

$$\text{where:} \quad w_{11}, w_{12}, w_{13} \cdots w_{1n} \text{ are weights in the Output layer}$$

$$
Error = \ RMSE(Openness, \widehat{Openness})
\tag{5.4}
$$

$$\text{where:} \quad \widehat{Openness} = predicted\ Openness$$

$$
Error = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Openness_i - \widehat{Openness_i})^2}
\tag{5.5}
$$

$$
\frac{\partial(Error)}{\partial w_{11}} = \frac{\partial}{\partial \widehat{Openness}} \left(\sqrt{\frac{1}{n} \sum_{i=1}^{n} (Openness_i - \widehat{Openness_i})^2} \right) \times \frac{\partial \widehat{Openness_i}}{\partial w_{11}}
\tag{5.6}
$$

5.1.1 Hyper-parameter Tuning

Hyper-parameter tuning is the process of selecting crucial parameters for a machine learning model. Tuning the neural network model involves selecting the network model parameters, these are:

- Activation Function

- Number of Neurons in Hidden Layer

- Batch Size

- Number of Epochs

Activation Function

The activation function is useful to model and capture the non-linear relationships between the dependent and independent variables. Several experiments were carried out with various activation functions: sigmoid, tanh, ReLU. RMSE values of neural network models with the various activation functions for the five personality traits are shown in Table 5.1. RMSE for tanh and sigmoid were low compared to ReLU activation function. Fit line plots are shown in Figure 5.2. Fit line plots for a regression model are scatter plots of actual values and the predicted values. The fit line or regression line should be 45 degrees to the x-axis. Here, the x-axis represents actual values ground truth and the y-axis represents predicted values. A regression model with more points close to the fit line is a better model. From the figure, it is apparent that neural networks with ReLU activation functions had more points close to the fit line than the sigmoid activation function. A similar trend was apparent with tanh activation function, fit line plots for other neural network models are presented in Appendix A. Based on these experiments, ReLU was used as the activation function.

Trait	Sigmoid	Tanh	ReLU
Openness	10 %	10.5 %	12.8 %
Conscientiousness	14.1 %	14 %	15.4%
Extraversion	17 %	17.2 %	20.6%
Agreeableness	12.7 %	13.6 %	15.1%
Neuroticism	18.2 %	13.6 %	22.4%

Table 5.1: RMSE values for different Activation Functions

(a) Openness, Activator = Sigmoid   **(b)** Openness, Activator= ReLU
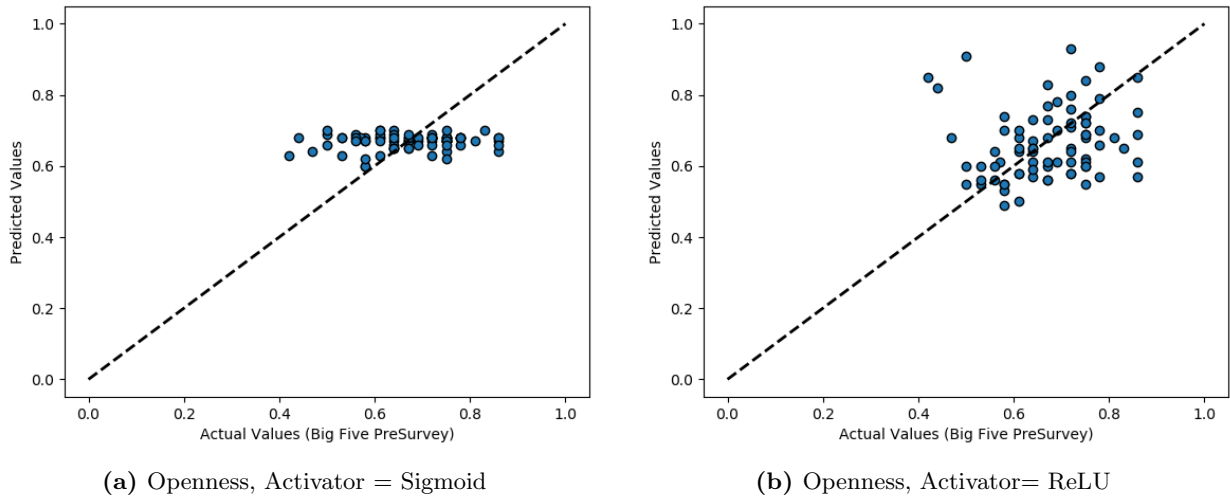
Figure 5.2: Fit Line Plots for Sigmoid and ReLU activation functions

Number of neurons in hidden layer

The number of neurons in hidden layer determines the extent of the non-linearity that can be captured from data. The number of neurons in the hidden layer is selected by running the neural network model with different numbers of neurons. Figure 5.3 shows the error convergence for different number of neurons in hidden layer. Here, the x-axis represents the number of neurons, and the y-axis represents the RMSE. At fewer than ten nodes in a hidden layer, the neural network was unable to capture the complex relationships in the data. As the number of neurons increases the model performance increases. With 30 neurons, the model produced acceptable results.
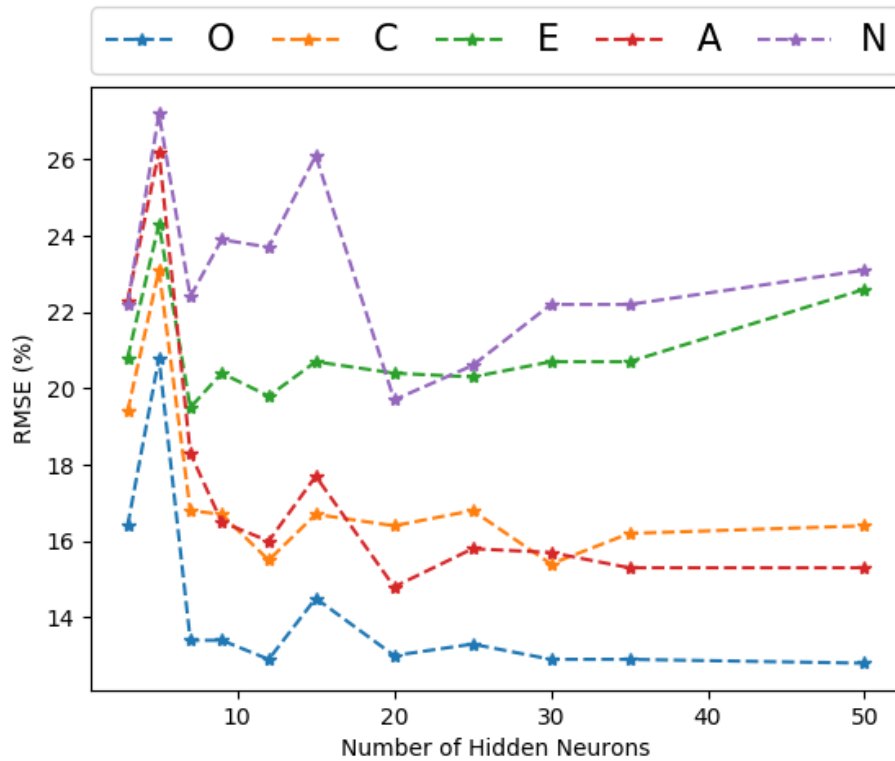
Figure 5.3: Convergence of error for Neural Networks based on neurons in Hidden Layer

Batch size

Batch size represents the number of training samples in one forward and backward pass. Figure 5.4 shows the error convergence of neural networks based on the batch size. Here, the x-axis represents the batch size and the y-axis represents RMSE. Root mean squared error decreases as batch size decreases because at higher batch sizes the model struggles to generalize for the wide range of included data. A batch size of 10 was employed (taking ten samples at a time to adjust the network parameters) because lower batch sizes can cause the network to become stuck in local minima [43]
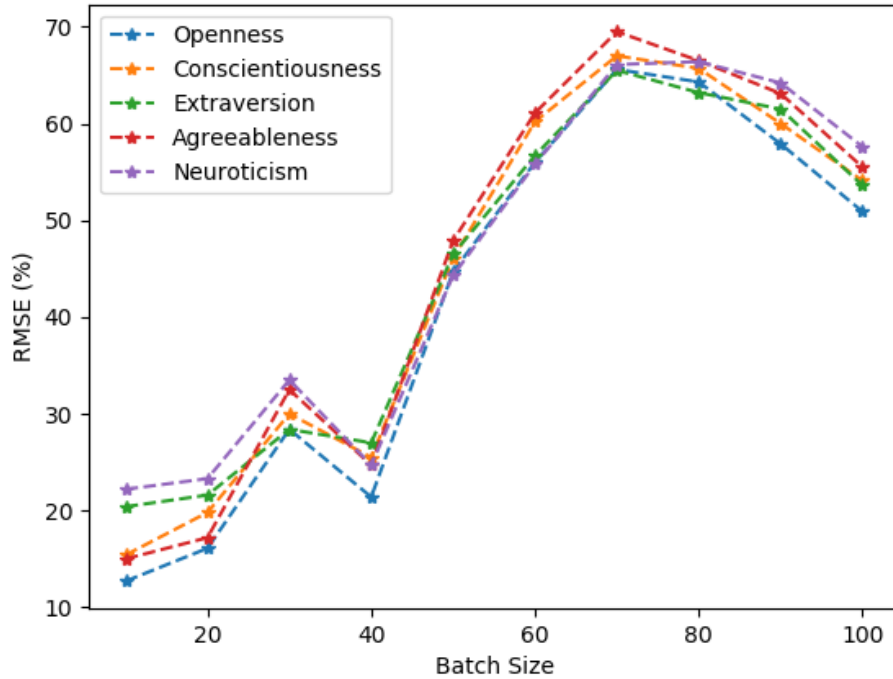
Figure 5.4: Convergence of error for Neural Networks based on training Batch Size

Number of Epochs

An epoch represents one forward pass and one backward pass over all the training samples. The number of Epochs represents the maximum number of iterations on the training data to learn and adjust the network parameters. Figure 5.5 shows the error convergence for different numbers of epochs. Here, the x-axis represents the number of epochs, and the y-axis represents RMSE. No improvements were observed.

Based on the baseline experiments, it was found that with 30 neurons, a batch size of 10 and 300 epochs produced acceptable results. After training the model, it was used to predict the Big Five traits of each participant on a daily basis. The average of all daily predictions for a participant was taken as the final predicted personality trait value of that participant. Each experiment was run 500 times, and the average of error values for 500 runs was taken as the final error value. RMSE value is presented as a percentage to aid interpretation.
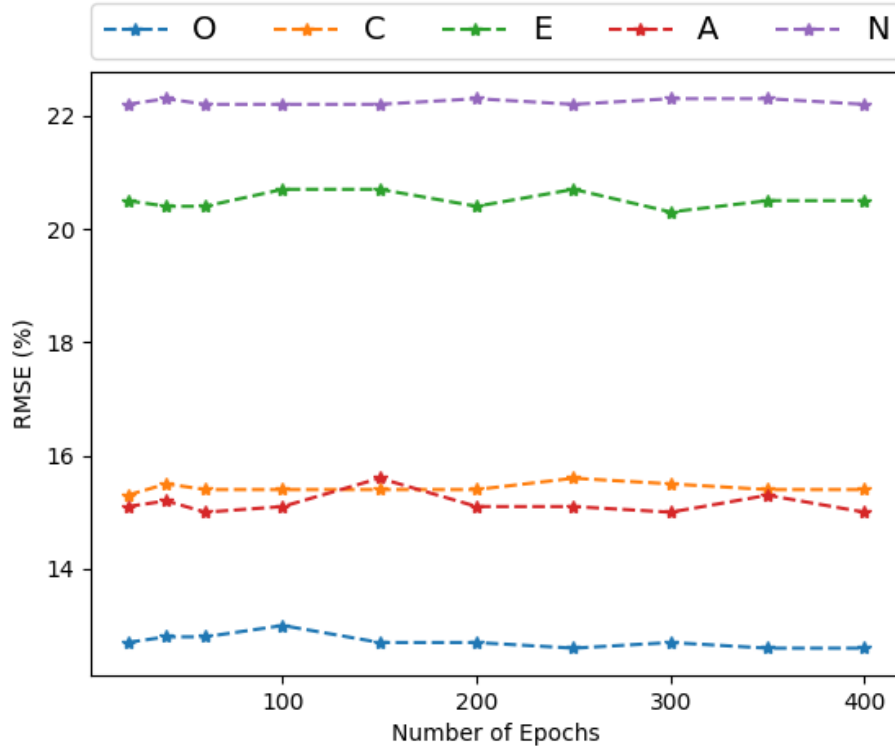
Figure 5.5: Convergence of error for Neural Networks based on Number of Epochs

5.2 Random Forests

5.2.1 Hyper-parameter Tuning

Hyper-parameters for the random forests model are:

- Number of Estimators

- Maximum Features

- Minimum Samples

- Maximum Depth

The number of estimators represents the number of decision trees constructed in a random forests model. For large datasets, a greater number of estimators help the random forests to generalize the data better because more decision trees are trained on different subsets of data. However, increasing the number of estimators can cause the model to overfit the data. Hence, an optimal number of estimators has to be chosen. Several experiments were carried out with different numbers of estimators, shown in Figure 5.6a. Here, the x-axis represents the number of decision trees, and the y-axis represents the RMSE. No significant improvements were noticed on increasing the number of estimators.

Maximum features denote the maximum number of features that can be used in constructing a decision tree. Figure 5.6b shows the error convergence for different numbers of features. Here, the x-axis represents the maximum features and the y-axis represents the RMSE. There was no improvement in the model performance for Openness, Conscientiousness, and Agreeableness. In the case of Neuroticism and Extraversion traits, error value decreases very slightly with the number of features but increases slightly after 5.

Minimum samples controls the minimum number of samples required for the algorithm to split an internal node while constructing the decision tree. As the minimum samples increases, the model has to consider a larger number of samples to split the node which might decrease the performance and results in underfitting. On the other hand, with a lower number of samples, the model is likely to overfit the data. Several experiments were carried out with different numbers of samples, and the results are shown in Figure 5.6c. Here, the x-axis represents the minimum samples and the y-axis represents RMSE. No significant improvements were noticed in the model.

Maximum depth represents the maximum depth each decision tree can grow in a random forests model. As the depth of the tree increases, the model can have more splits to capture the information in the data, this is recommended for larger datasets because for small datasets this can result in overfitting. Figure 5.6d shows the error convergence for various depths. Here, the x-axis represents the maximum depth and the y-axis represents the RMSE. For all models except the Openness trait, slight improvements can be noted in model performance with increasing the depth, but after reaching 20, no improvements were evident.

Tuning the hyper-parameters for random forests did not provide a significant improvement in the model performance (in terms of RMSE) other than few minor gains. Final experiments were carried out with five decision trees and a maximum depth of 20. Maximum features and minimum number of samples required to split a node were 5 and 50 respectively. Mean square error was used for split criteria. Models were used to predict the Big Five traits of each participant on a per day basis, and the average of all daily predictions for the participant was taken as the final predicted personality trait. Each experiment was run 500 times, and the average of error values for 500 runs was taken as the final error value.
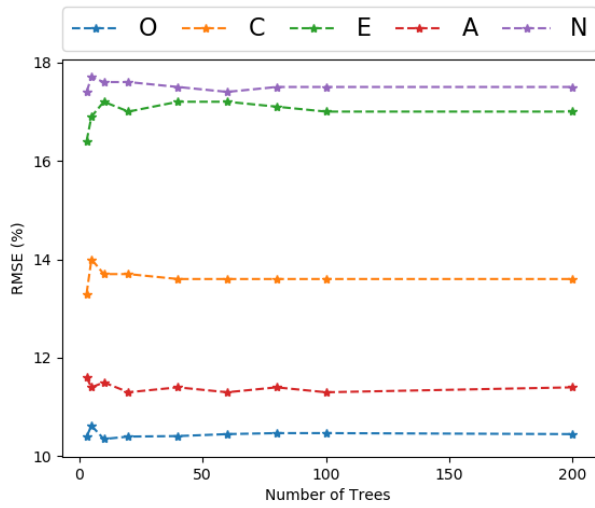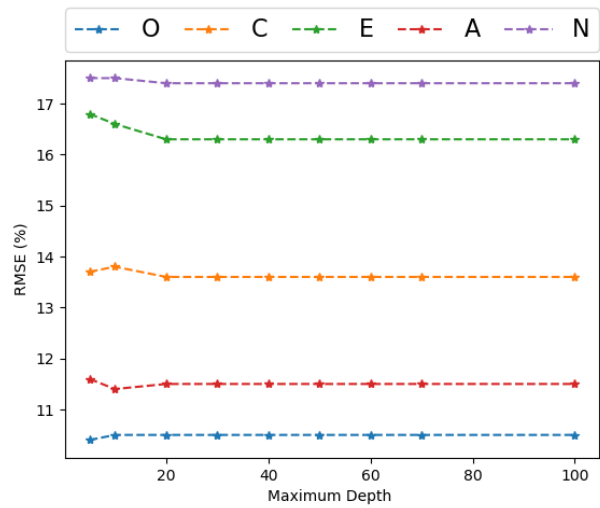
(a) Error Convergence: Number of Estimators
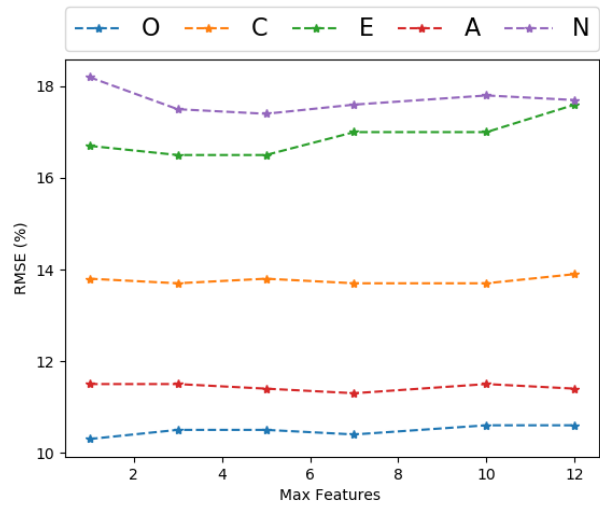
(b) Error Convergence: Maximum Features
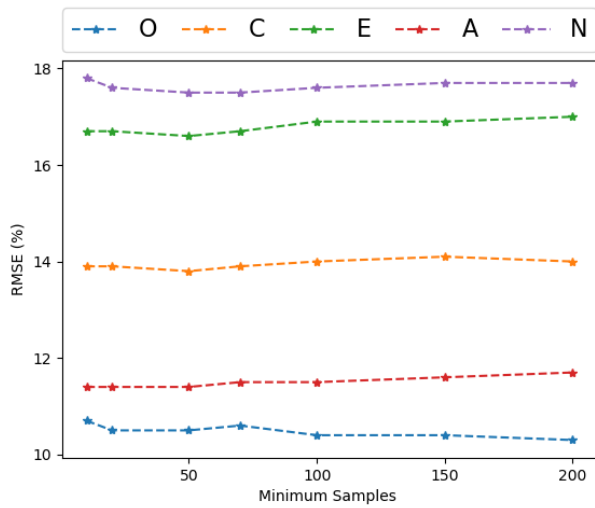
(c) Error Convergence: Minimum Samples

(d) Error Convergence: Maximum Depth

Figure 5.6: Tuning Random Forests models

5.3 Support Vector Machines (SVM)

Regression modeling problems can be converted into classification modeling problems. In this work, the Big Five personality, a regression modeling problem was converted into a classification modeling problem. The purpose of this experiment is to establish the guidelines in reporting the results while using classification learning models for behavioral studies. Personality trait values were converted into *low ('L')* or *high ('H')* based a mean value split. For example, in case of Neuroticism trait mean value of the population was 0.48, hence participants with less than 0.48 Neuroticism trait value were labeled as *low ('L')* and rest were labelled as *high ('H')*. Table 5.2 shows a sample of 5 records after labeling the data for classification modeling. Here, the left side columns are Big Five trait values for individuals and the right-hand side are the new class labels. After converting all the data, class distributions were almost equally distributed across the class labels for all five traits as shown in Table 5.3, this ensured that there was no skew or class bias.

ID	E	A	C	N	O		E	A	C	N	O
						Mean:	0.6	0.8	0.7	0.48	0.67
506	0.78	0.97	0.83	0.14	0.75		H	H	H	L	H
514	0.53	0.56	0.69	0.54	0.53		L	L	L	H	L
519	0.59	0.81	0.69	0.5	0.75		L	H	L	H	H
532	0.81	0.81	0.69	0.61	0.72		H	H	L	H	H
534	0.91	0.72	0.83	0.71	0.86		H	L	H	H	H

Table 5.2: Binned Labels based on the mean of the population trait

E- Extraversion, A- Agreeableness, C- Conscientiousness, N- Neuroticism, O- Openness

Trait	Low	High
Openness	52%	48%
Conscientiousness	54%	46%
Extraversion	55%	45%
Agreeableness	45%	55%
Neuroticism	47%	53%

Table 5.3: Class Distribution Statistics after binning

5.3.1 Hyper-parameter Tuning

Hyper-parameters for SVMs are:

- Kernel

- Gamma

- C

The Kernel enables the SVM to capture the non-linearity presented in the data by transforming the data to higher dimensions. Different kernels use different strategies to transform the data. For instance, the *rbf* kernel adds similarity features and *poly* kernel adds polynomial features. Several experiments were carried out with three popular kernels namely *rbf, poly, sigmoid* and the accuracies are shown in Table 5.4. For Openness, Agreeableness and Neuroticism the *poly* kernel resulted in better accuracies. The *Rbf* kernel performed well for Conscientiousness and Extraversion traits.

Trait	rbf	poly	sigmoid
Openness	52.5%	62.5%	51.2%
Conscientiousness	53.7%	48.7%	43.7%
Extraversion	55%	48.7%	55%
Agreeableness	56.2%	58.7%	47.5%
Neuroticism	52.5%	55%	55%

Table 5.4: Accuracies for different Kernels

Gamma, also known as the kernel coefficient, is useful to control the amount/range of influence for existing data samples on the newly transformed space. At higher gamma values, each data instance has a low influence on the new features. A small gamma allows the data instances to have a large range of influence on the transformed data and makes the margin or decision boundary smoother. Figure 5.7a shows the accuracies of SVM models for various gamma values. Here, the x-axis represents gamma values and the y-axis represents accuracy. For Openness, Conscientiousness and Neuroticism traits, models perform better at lower gamma values. There was no change in accuracy for Agreeableness trait.

Sometimes it is difficult for the SVM model to linearly separate the entire data and find hard margin between the classes, in such cases margin violations can be allowed using the C as a penalty parameter to the model. Several experiments were carried out with different C values, and the results are shown in Figure 5.7b. Here, the x-axis represents the C values and the y-axis represents accuracy. At higher C values the models try to find small margins between the classes allowing fewer margin violations, resulting in many mis-classifications. Conversely, lower C value allows significant margin violations and finds boundaries between the classes resulting in improved accuracy. In the case of Openness, Agreeableness and Neuroticism

(a) Accuracy for various Gamma values

(b) Accuracy for various C values
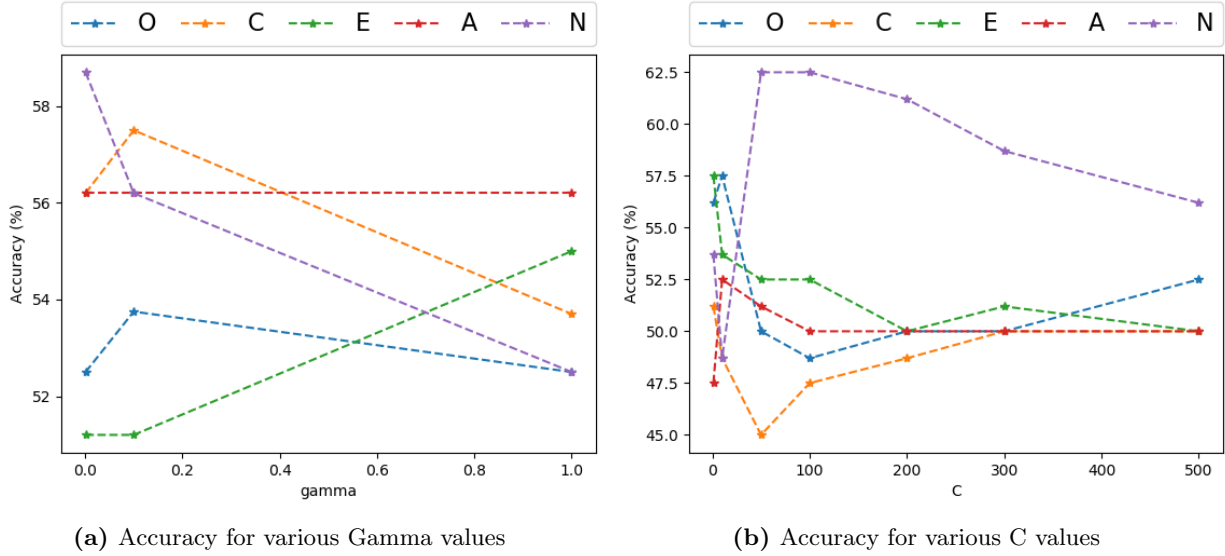
Figure 5.7: Tuning SVM models

models had slightly low accuracies at lower C values, and accuracy increases as C increases to a maximum. Models for Conscientiousness and Extraversion produced good results at lower C values.

SVM models for all five traits personality traits performed well with different sets of parameters, these parameters are presented in Table 5.5. Models were trained using the selected parameters and used to predict the Big Five traits of each participant on a daily basis, and then the most common class label of all daily predictions of a participant was taken as the final predicted personality trait value of that participant. Each experiment was run 500 times, and the average of accuracy for 500 runs was taken as the final accuracy.

Trait	kernel	C	gamma
Openness	poly (degree - 3)	50	0.01
Conscientiousness	rbf	1	0.01
Extraversion	rbf	1	0.01
Agreeableness	poly (degree - 3)	50	0.1
Neuroticism	poly (degree - 3)	50	0.01

Table 5.5: Hyper-Parameters for Personality Traits

Chapter 6

Results

6.1 Regression Models

To analyze, compare, and illustrate the predictive power, and the quality of regression models, statistical, fit line and residual analysis were performed. Statistical analysis indicated that random forests produced good results. However, fit line and residual analysis illustrated that neural network models were more capable of generalizing the data. Full details are mentioned in the following sections.

6.1.1 Statistical Analysis (RMSE)

Regression models were compared using RMSE values. Figure 6.1 shows the RMSE values for Neural Networks, Random Forests and Mean Models. Here, the x-axis represents the root mean squared error and the y-axis represents the Big Five personality traits. Error values are presented in percentage for ease of interpretation. RMSE values for random forests and mean models are same for Openness and Extraversion traits. For Conscientiousness, Agreeableness, and Neuroticism, random forests models performed slightly better than mean models. Error values of neural networks for Extraversion and Neuroticism traits are more than 3 % better compared to mean models. In case of Agreeableness and Neuroticism, neural network RMSE values are around 4% more than the values for the Random Forests model. Overall, in terms of statistical metric RMSE, neural networks models for all five personality traits were outperformed by random forests and mean models. Random forests models performed better than mean models and neural networks.

6.1.2 Fit Line Plots Analysis

For fit line plots, on the x-axis, actual personality traits values (ground truth) are plotted, on the y-axis values predicted by the machine learning models are plotted. Fit line plots for mean models, random forests models and neural networks models are presented in Figure 6.2, Figure 6.3 and Figure 6.4, respectively. In all fit line plots, the x-axis represents the actual personality trait, the y-axis represents the predicted values from machine learning models, and the dotted diagonal line represents the regression line. A machine learning model that has many points close to the regression line is best. Figure 6.2 shows the fit line plot of Mean Models for all Big Five Personality traits. Here plots for all five traits are shown in a single image. Since
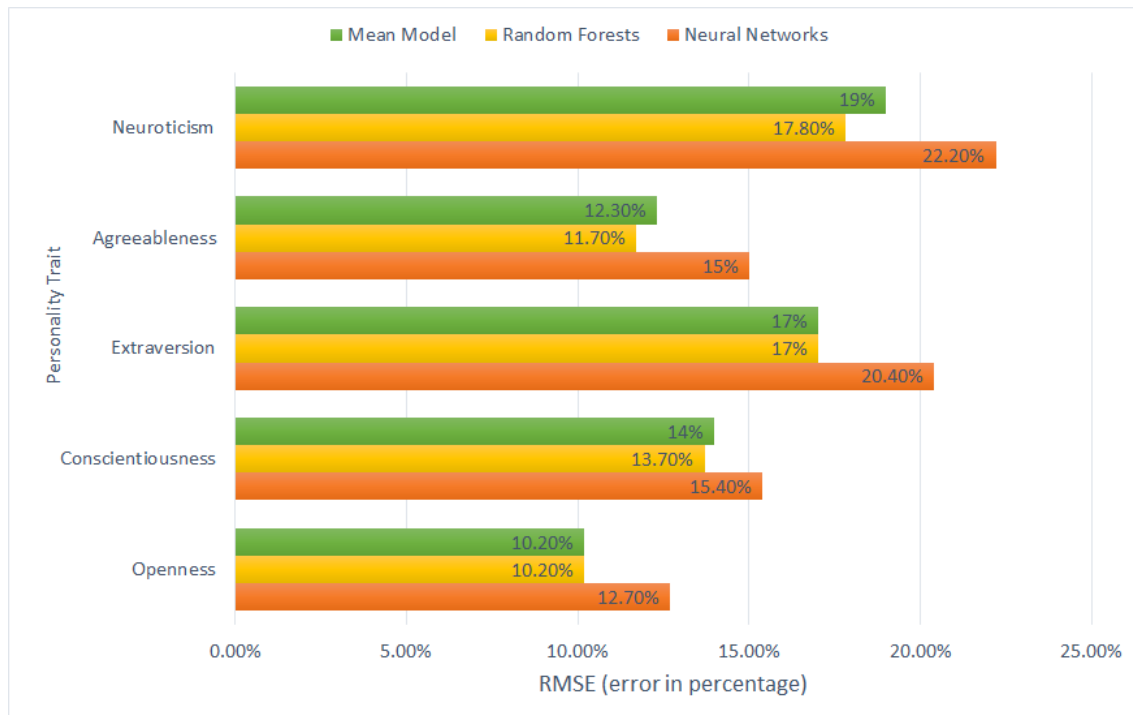
Figure 6.1: Root Mean Square Errors for Regression Models

these are mean models all the points are distributed as flat lines across the fit line, it can be clearly noticed that there are not many points close to the fit line, indicating that baseline models generalized poorly.

Figure 6.3 shows the fit line plots for random forests models. Each Big Five personality trait has its own plot. Compared to baseline models there are more points close to the fit line for all the five traits. Overall, random forests models performed better than the baseline models. However, points are distributed similarly to baseline models. This could be because the model is failing to generalize the data and randomly guessing around the mean of the population.

Fit line plots for neural network models are shown in Figure 6.4, each Big Five personality trait has its own plot. The model for Openness trait (Sub Figure 6.4a) has many points close to the fit line with a few outliers present. Similarly, Conscientiousness and Agreeableness traits also have many points close to fit line or regression line but with more outliers. Among all five traits, Neuroticism performed poorly with few points close to fit line and many incorrect predictions. Similar trends were observed for the Extraversion trait. Among all regression models, neural network models have many points close to the regression line with outliers, but for baseline models and random forests models, there are fewer than 10% of the points closer to the regression line. It is clearly evident that Neural Network models attempt to make a prediction along the line, whereas the baseline model and random forests abandon the line entirely. This implies that neural network models are making more meaningful predictions than the remaining models.
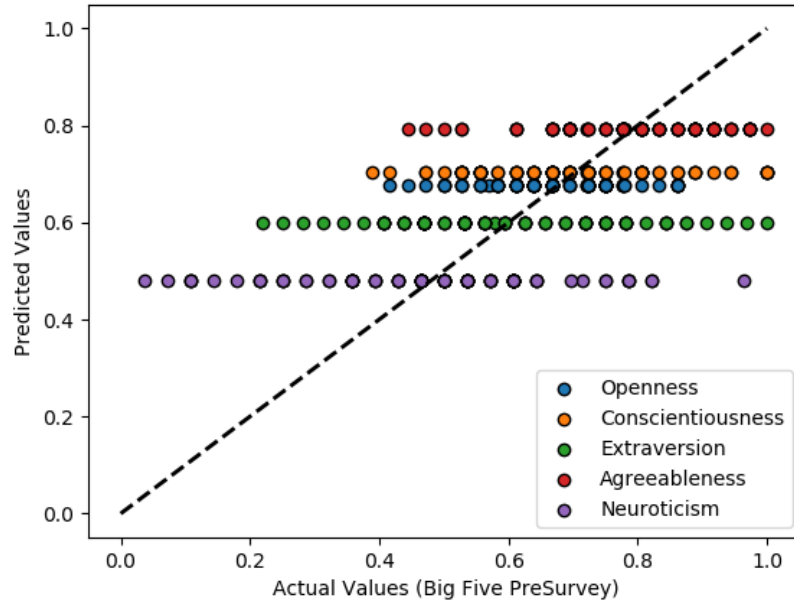
Figure 6.2: Fit Line Plots for Mean Models

6.1.3 Residual Plots Analysis

In residual plots, actual values versus residual values are analyzed. A Residual value is difference between the predicted value and the actual value. A horizontal line is also presented in residual plots representing performance line where error/residual is zero. The distance for each point from this line indicates the error or offset from the actual value for that point. Residual plots for mean models, random forests and Neural Network models are shown in Figure 6.5, Figure 6.6 and Figure 6.7. In all residual plots, the x-axis represents the actual values, and the y-axis represents the residual values. For residual plots, baseline models have monotonically increasing residuals with the independent parameter, indicating an exceptionally poor model, characterized by strong systemic error. A similar trend is noted with random forests models. On the other hand, neural network models have random residuals indicating that the model is better at representing the phenomenon based on independent variables but susceptible to noise.

Looking just at the error values might suggest that the baseline model provided better performance than the neural network model; this result is largely an artifact, as most participants are close to the mean, guessing the mean leads to apparent accuracy. A closer look at the fit line plots and residual plots indicated that neural networks models outperformed baseline models and random forests models. This is the first analysis in behavioral studies to report accuracy with respect to a baseline model, in terms of statistical metrics, fit line plots and residual plots. It is worth noting that while the results are close to the baseline estimates in statistical comparison, they are substantially better than the closest error values in existing research. It can be speculated that such previous work would fare even worse in comparison to a baseline model.
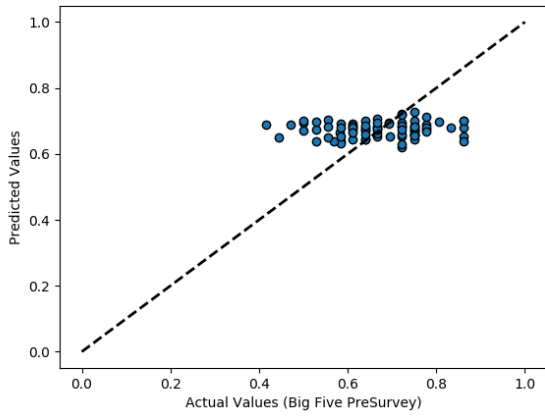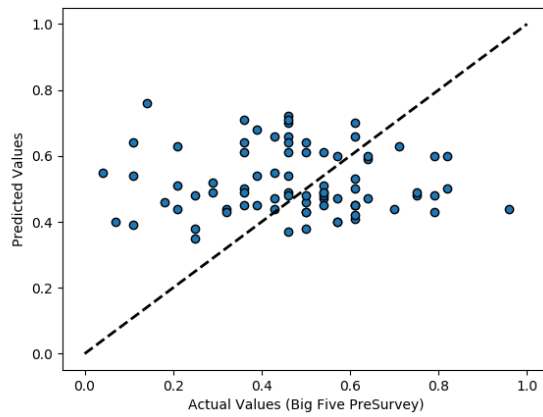
(a) Openness
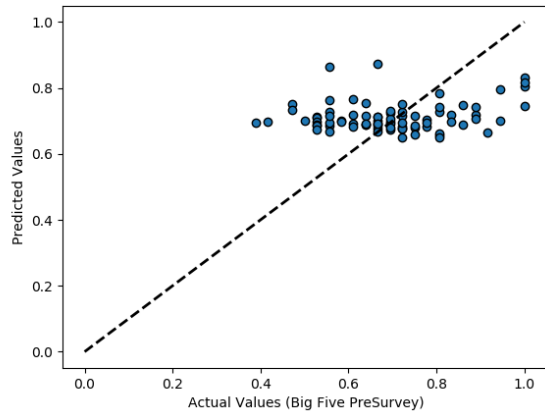
(b) Conscientiousness
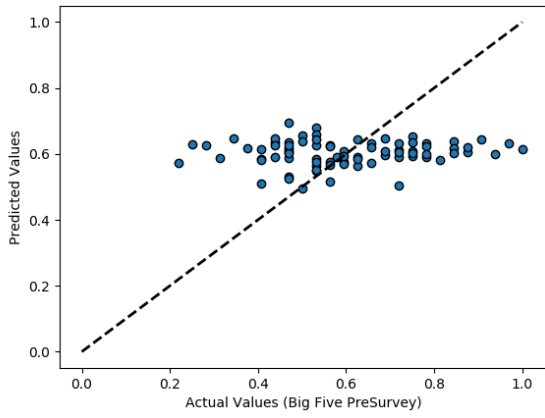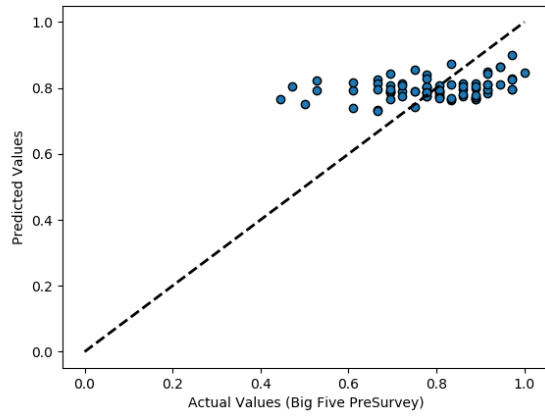
(c) Extraversion
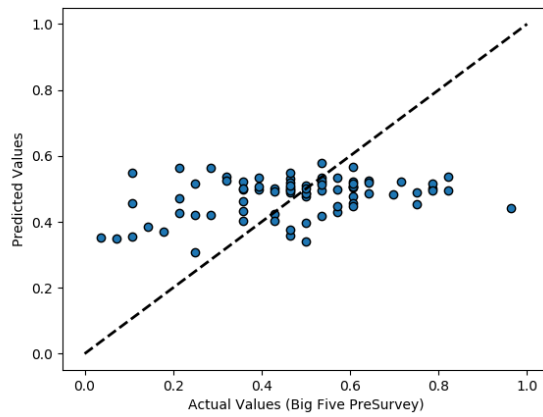
(d) Agreeableness

(e) Neuroticism

Figure 6.3: Fit Line plots for Random Forests Models
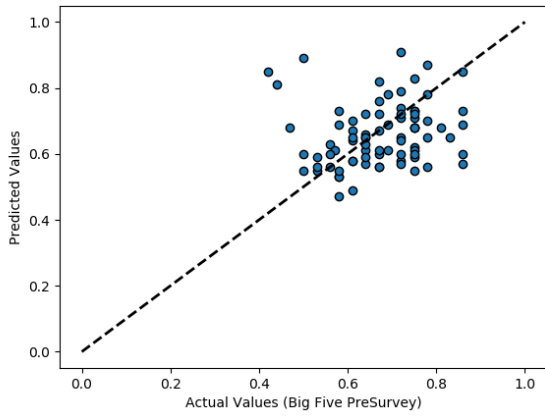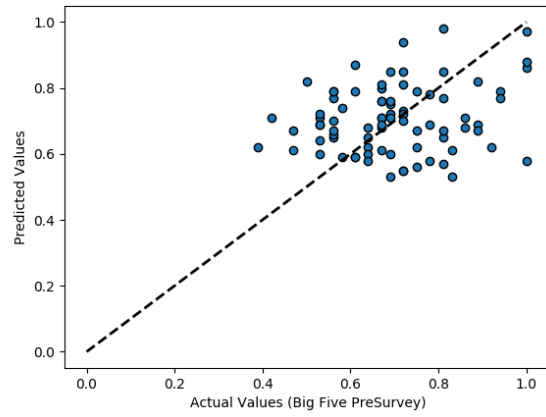
(a) Openness

(b) Conscientiousness
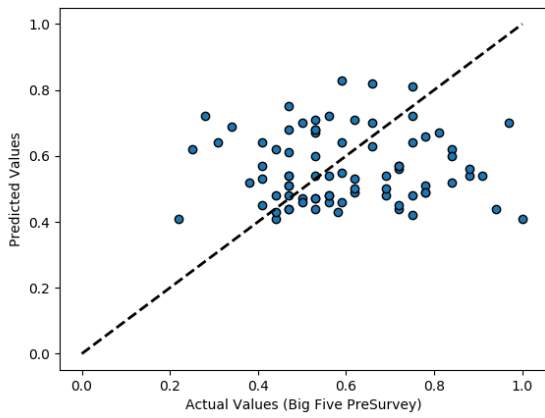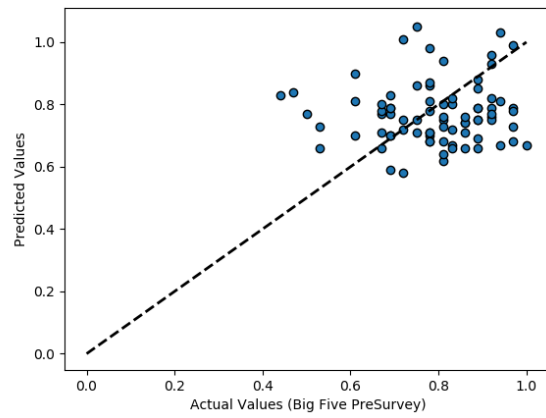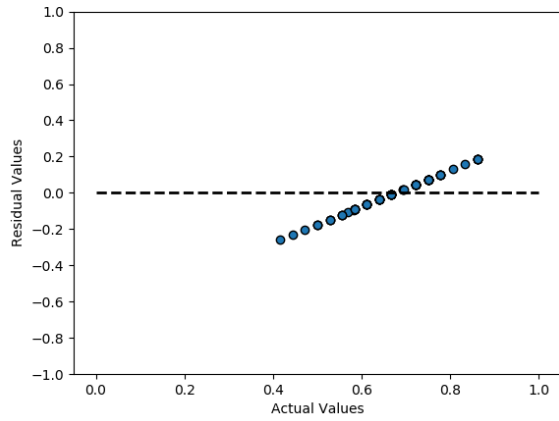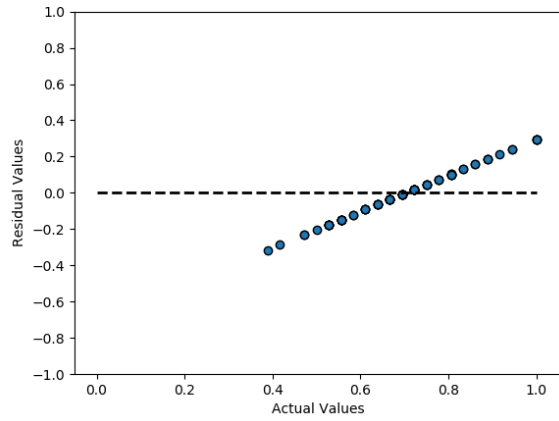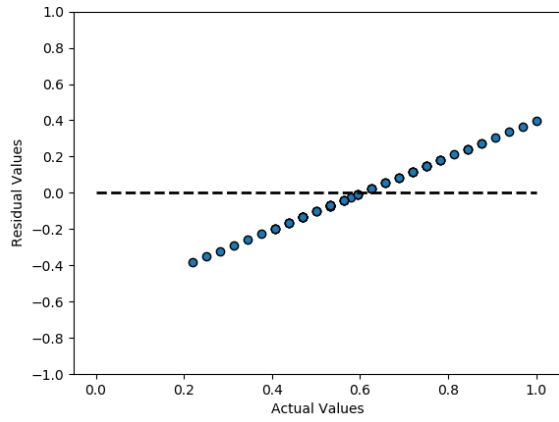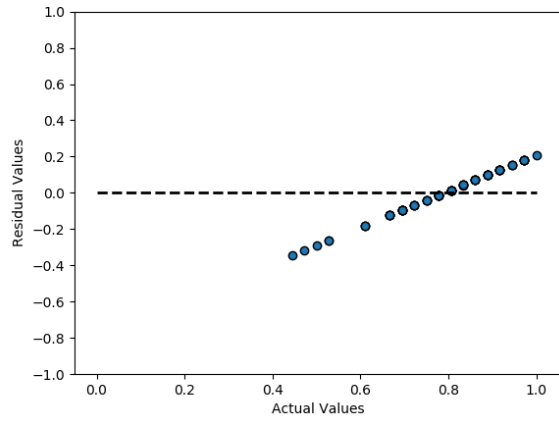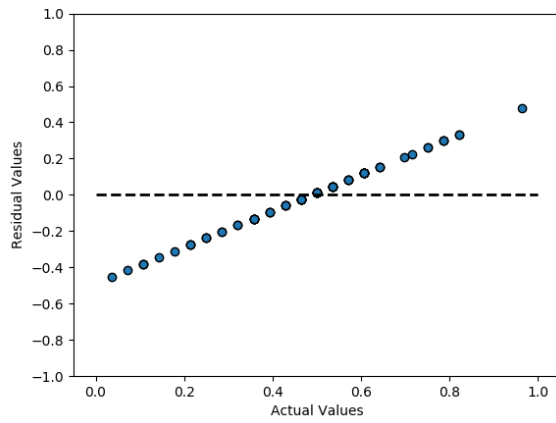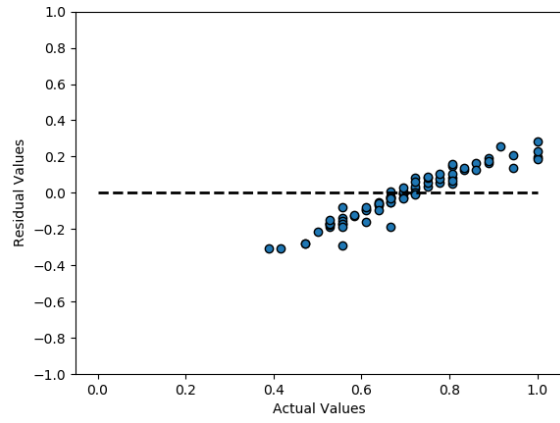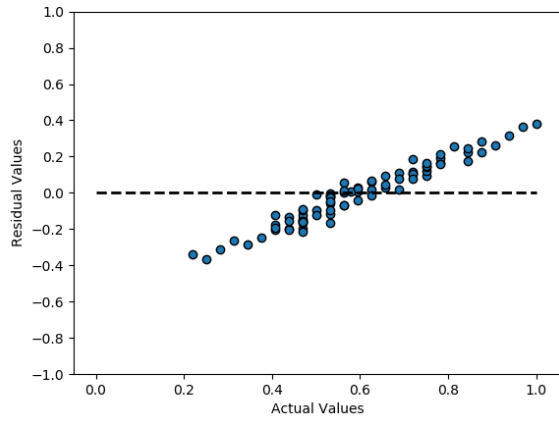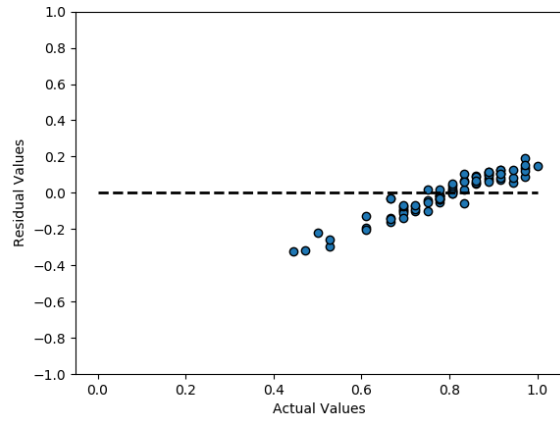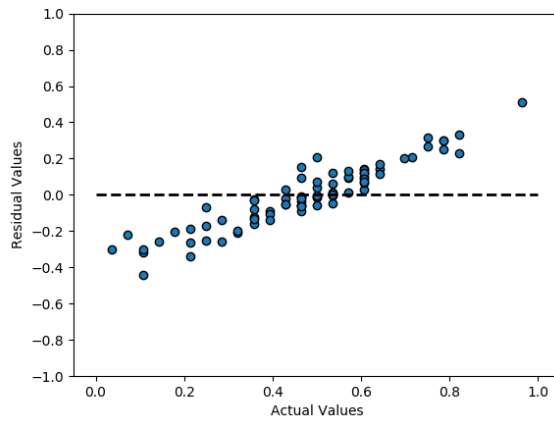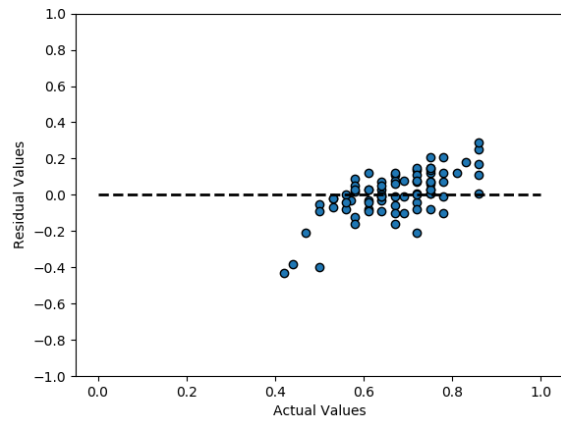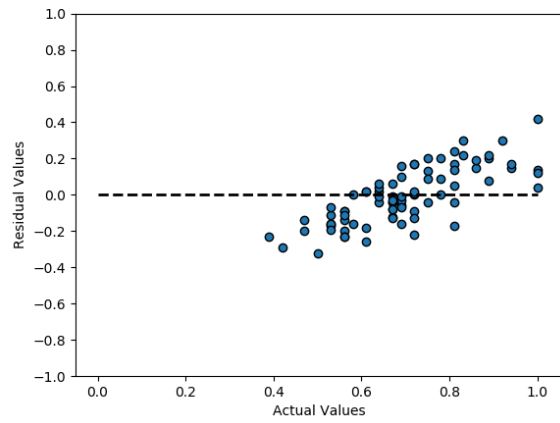
(c) Extraversion

(d) Agreeableness

(e) Neuroticism

Figure 6.4: Fit Line plots for Neural Network models

(a) Openness

(b) Conscientiousness

(c) Extraversion

(d) Agreeableness

(e) Neuroticism

Figure 6.5: Residual plots for Baseline Models
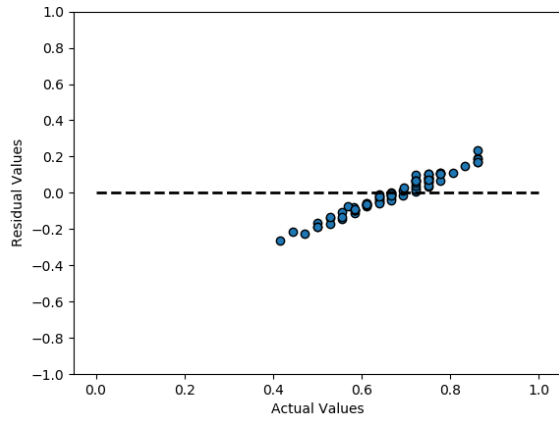
(a) Openness

(b) Conscientiousness

(c) Extraversion

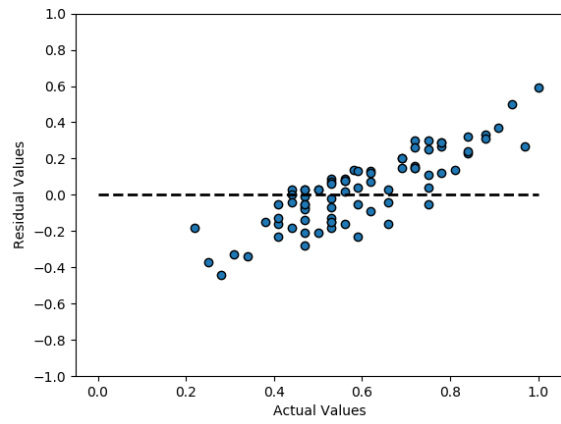**(d)** Agreeableness

(e) Neuroticism

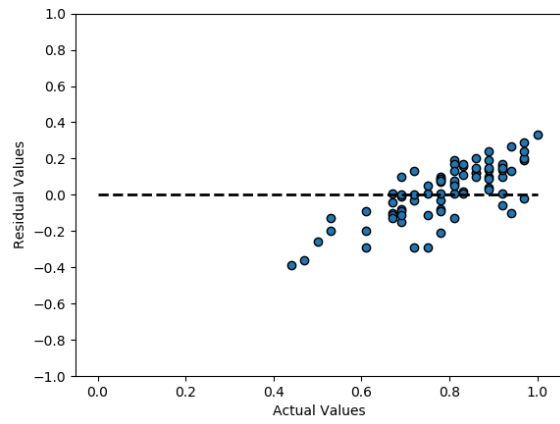**Figure 6.6:** Residual plots for Random Forests Models
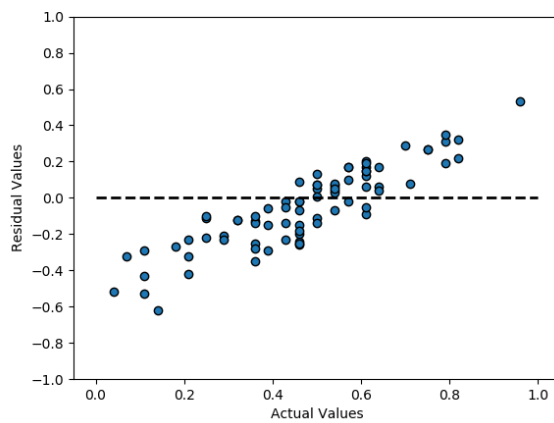
(a) Openness

(b) Conscientiousness

(c) Extraversion

(d) Agreeableness

(e) Neuroticism

Figure 6.7: Residual Plots for Neural Network models

6.1.4   Neural Networks Error Range Analysis

As neural networks models made more meaningful predictions, neural networks model error ranges are presented in detail in this section. Neural network models predicted the Big Five personality traits Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism with root mean squared errors of 12.7%, 15.4%, 20.4%, 15%, 22.2% respectively when compared to pre-questionnaire results. These results are substantially better than previous studies [24, 30, 90, 125]. Among the Big Five traits, neural network models predicted Openness, Conscientiousness, and Agreeableness with lower error values (and low enough for us to have confidence in the prediction) compared to Extraversion and Neuroticism. Though the error values were high enough to make useful prediction unlikely for these two traits, they are still marginally better than the existing studies. Figure 6.8 shows the number of participants in each error bin for the Big Five personality traits, here x-axis represents the error bins (0-5 %, 5-10 %, 10-20 %, 20-30 %, 30-45 % and 45-65 %) and y-axis represents the number of participants predicted by Neural Networks models with the specified error bin range.

Neural Networks predicted the Openness trait of 30 participants with less than a 5% error value and 23 participants with an error value ranging from 5-10 %. From a total of 80 participants, 66% participants' Openness trait was predicted with less than 10% error. The Conscientiousness trait for 81% of the population was predicted with less than 20% error. Similarly, the model for Agreeableness trait predicted 85% of the population with less than 20% error value. Extroversion and Neuroticism models predicted the traits for approximately 65% of the population with less than 20% error. Extraversion and Neuroticism both had errors in excess of 30% for 11 out of 80 participants.
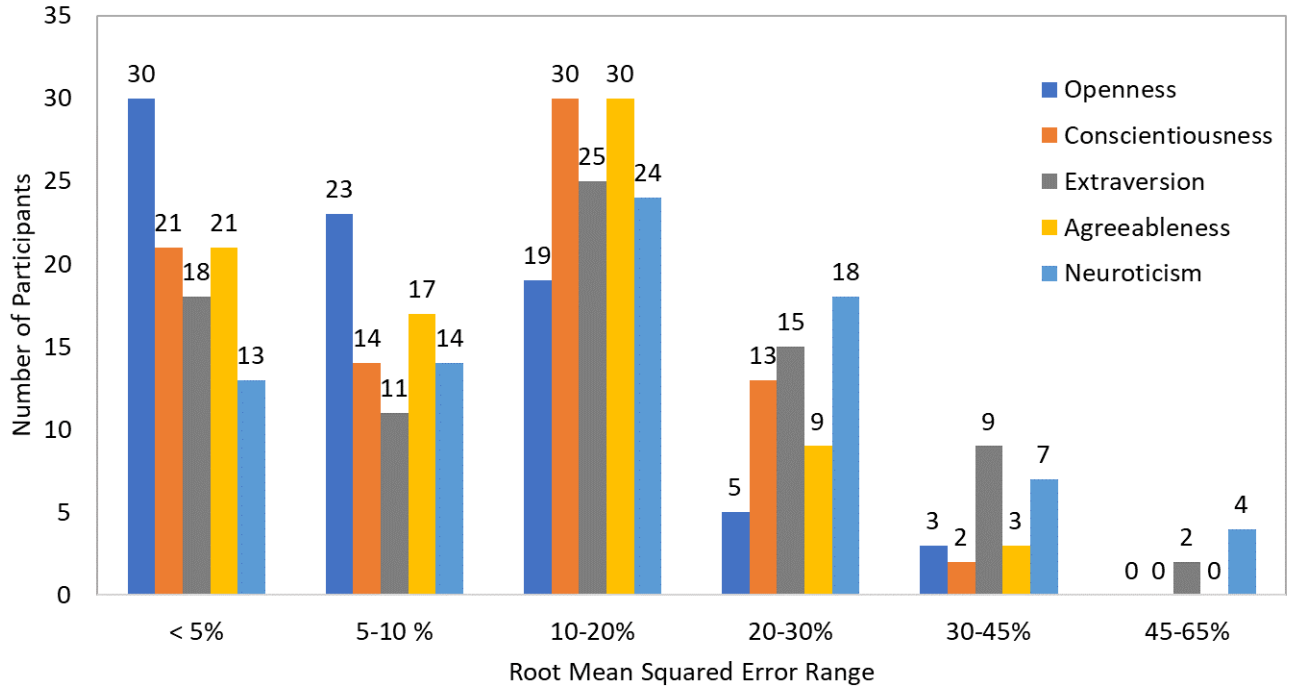
Figure 6.8: Number of Participants based on Error Range, total participants=80

6.2 Classification Models

In order to compare and illustrate the prediction of the classification models, statistical and confusion matrix analysis was performed. Full details are mentioned in the following sections.

6.2.1 Statistical Analysis (Accuracy)

The accuracy for baseline models and SVM models are shown in Figure 6.9. Here, the x-axis represents the accuracy in percentage, and the y-axis represents the personality trait. SVM models performed better than baseline models for Openness and Neuroticism traits. In the case of Agreeableness and Extraversion, the performance improvement is marginal. The SVM model for Conscientiousness trait was no better than the baseline model. In terms of statistical metric comparison, it appears that baseline models and SVM models have similar accuracy for Conscientiousness, Extraversion, and Agreeableness. To analyze the models further confusion matrices for baseline models and SVM models are presented.

6.2.2 Confusion Matrix Analysis

A confusion matrix is a matrix representation of true class labels and the corresponding predicted class labels. Confusion matrices are helpful in understanding how well a model can generalize the data. In a confusion matrix, the row represents the total number of true class labels present in the data, and the column represents
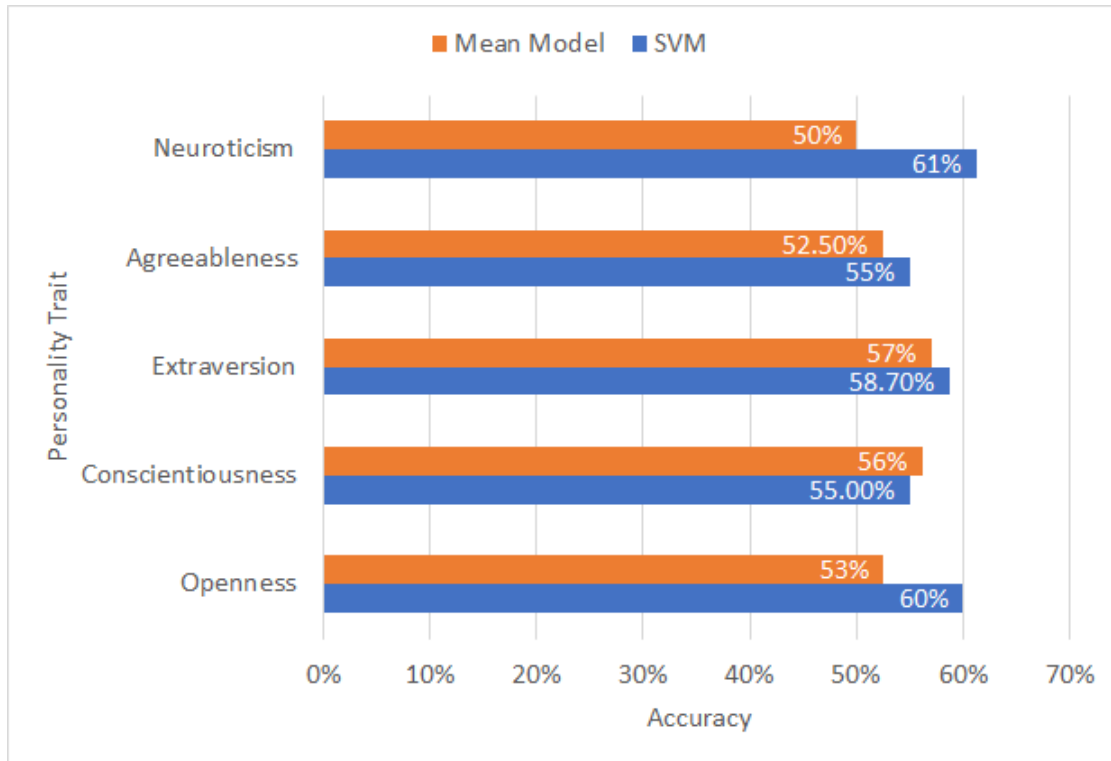
**Figure 6.9:** Accuracies for Classification Models

the total number of instances model classified as that class. The diagonal elements represent the number of instances correctly predicted. Models with a confusion matrix having more values on the diagonal indicate a better representation of the data.

Confusion matrices for baseline models are shown in Figure 6.10. Confusion matrices for Openness, Conscientiousness, and Extraversion indicate that for all data points having *Low* and *High* as true class labels, baseline models predicted the class label as *Low* class. This is expected since a majority of the population has *Low* as class label. A similar trend can be noticed for Agreeableness and Neuroticism, but with *High* class label since a majority of the population has *High* as class label.

Confusion matrices for SVM models are presented in Figure 6.11. For the Openness trait, 71% of the data points were correctly predicted as the *Low* class and 29% data points are wrongly classified as *High* class. Similarly among all data points labeled as *High* class, 47% were correctly predicted as *High* and remaining are classified as *low*. The Conscientiousness model clearly struggled to classify the *High* class labeled samples and wrongly predicted 74% of them as *Low*, only 26% *High* class samples were correctly predicted. On the other hand model could correctly classify 78% of *Low* class samples. For Extraversion trait 70% of *Low* class samples are correctly classified and 44% of *High* class labeled samples were correctly classified. The Agreeableness model correctly classified 45% of *Low* class samples and 64% of *High* class samples. In case of Neuroticism trait 88% of the *High* class samples were correctly classified, but 65% of the *Low* class samples were incorrectly classified.

Though it appears that baseline models performed on par with SVM models for Conscientiousness, Extraversion, and Agreeableness, confusion matrices indicated that baseline models underperformed SVMs. Looking at the accuracies of SVM models for all five traits, it appeared that the Neuroticism trait was best predicted among the five traits with 61% accuracy, but further analysis of confusion matrices indicated that Neuroticism model performed poorly.

Overall, model accuracies are low compared to previous studies [24, 115, 124]. SVM models clearly struggled to generalize the data. This shortcoming is likely due to several reasons. First and foremost, as noticed in the correlation study, data may not be linearly separable implying that an SVM model is unlikely to perform well. Second, the amount of time and effort spent on tuning the SVM models were low compared to neural network models. Additional parameter tuning may have improved the results. The primary purpose of the classification experiment was to establish the guidelines in presenting the results for behavioral studies, and the SVM models were employed for comparison with prior work. Finally, model generalizability was further weakened by kernels which might have hindered generalization, thus increasing the already struggling model's difficulties to find linearly separable vectors.
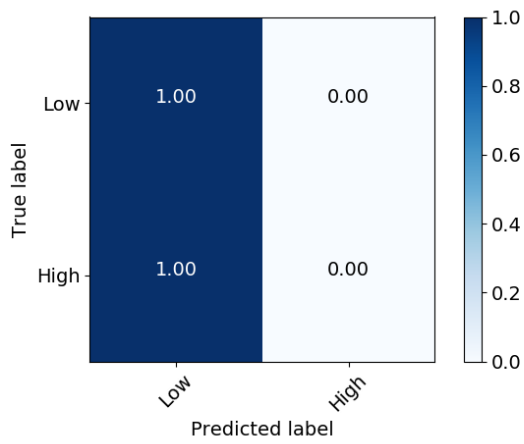
(a) Openness
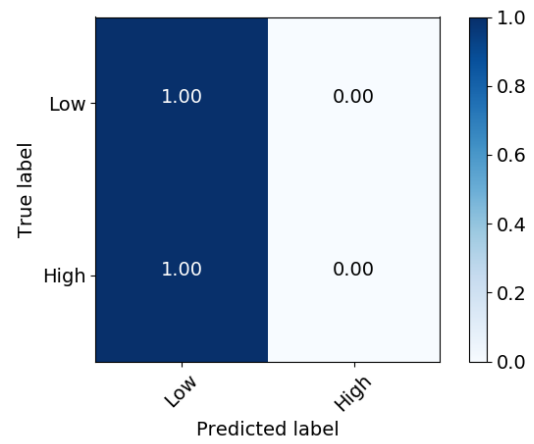
(b) Conscientiousness
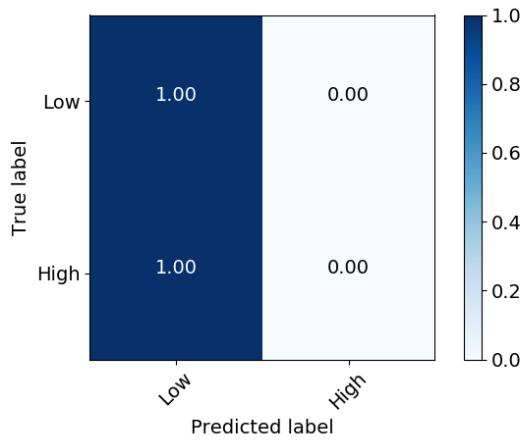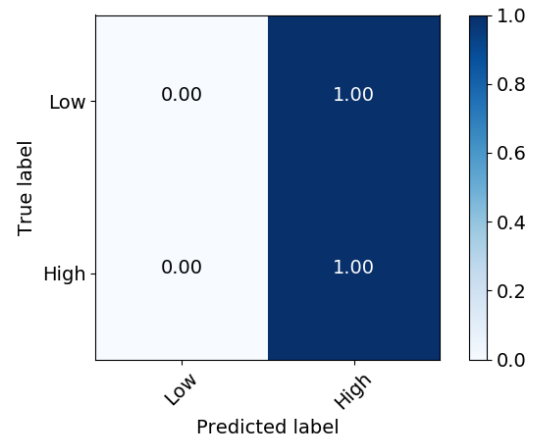
(c) Extraversion
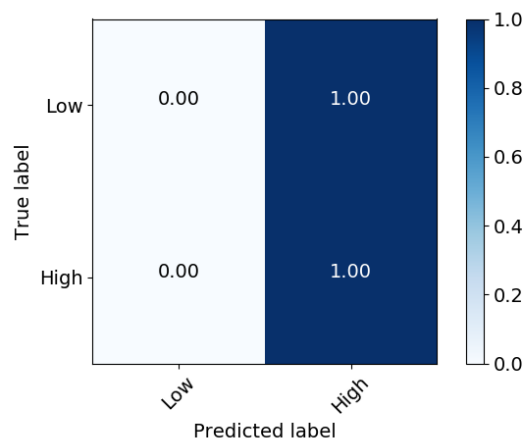
(d) Agreeableness

(e) Neuroticism

Figure 6.10: Normalized Confusion Matrices for Baseline Models

(a) Openness

(b) Conscientiousness

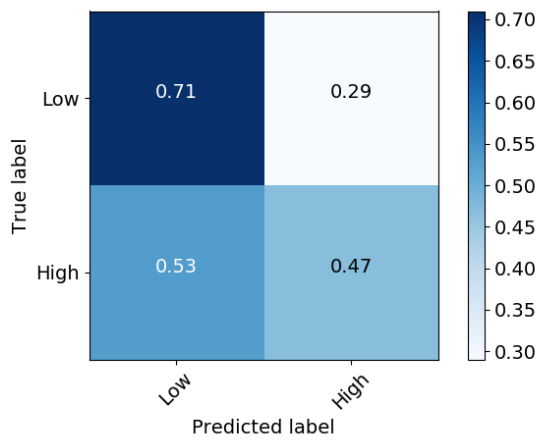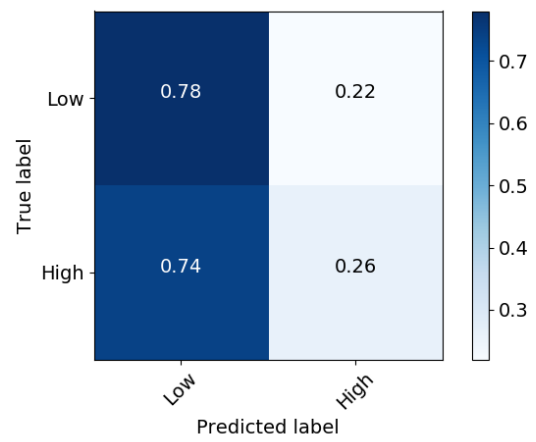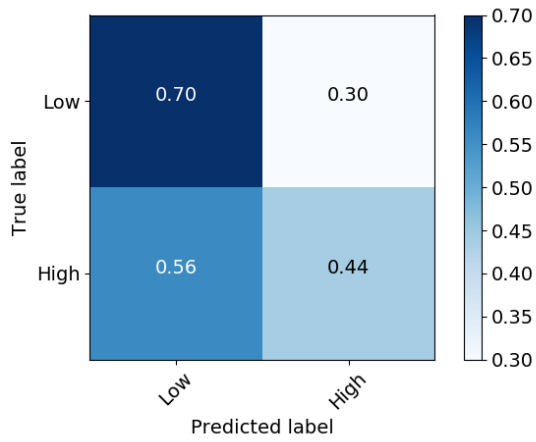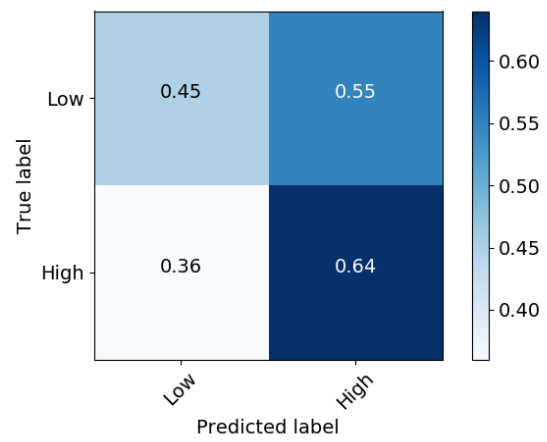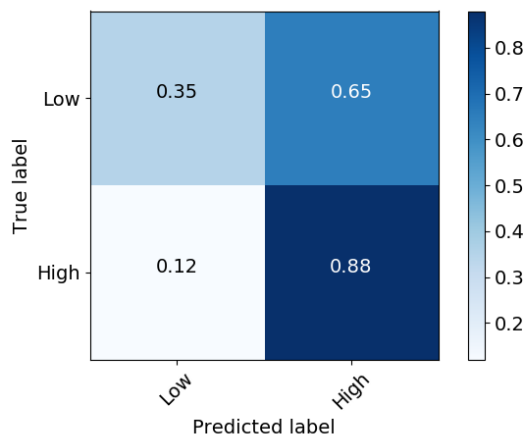**(c)** Extraversion

(d) Agreeableness

(e) Neuroticism

Figure 6.11: Normalized Confusion Matrices for SVM Models

Chapter 7

Discussion & Future Work

In this work, machine learning models which provide the ability to predict the personality based on real-world mobility, habits, and communications behavior were created. The models, based on a neural network architecture were able to approximate participant personality with an average error value of 12.7-22.2%, substantially outperforming most prior work. These findings were based on not only app usage data, as has been previously reported but behavioral parameters (such as charging time) and geospatial behavior parameters (such as anchor point arrival time). That personality correlates with behavior, and that geographic parameters can improve these models is an important secondary contribution of this work and a potential guidepost for other researchers in this area. This is the first work to successfully model personality across all five dimensions along a continuum, as a regression rather than classification problem, using readily available smartphone sensor data. Finally, results were compared to a baseline algorithm. Error values were comparable but systemically lower than the baseline. This is an important step in establishing rigor in reporting the reliability of estimates.

Neural network models created in this work predicted the Big Five personality traits Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism with root mean squared errors of 12.7%, 15.4%, 20.4%, 15%, 22.2% respectively when compared to pre-questionnaire results. In the literature, a majority of the studies that have modeled the personality as a Classification problem reported the accuracies in the range of 50-83% when the personality is binned into two classes [2, 24, 124]. These classification studies have several limitations. First, model performance is impacted by the number of bins the personality is put into. Studies that have binned the personality into three classes reported the accuracies in the range of 50-70% [30, 115, 124, 125] whereas studies that have used two bins achieved a higher accuracy of 83%. Second, several studies have put the personality traits into two bins [low, high] which may not be very useful in distinguishing people whereas regression modeling provides a more accurate way to do the same.

In case of the few studies that have attempted to model the personality as a regression problem, the best performing study reported the root mean squared errors around 28% [8]. Two regression studies have reported the mean absolute error values to be around 11%, but these errors cannot be directly compared to root mean squared errors [45, 46]. The problem with mean absolute error is that negative errors cancel out the positive errors thus returning low error values whereas the root mean squared error squares the error thus preserving the magnitude of the error [22]. Overall, Neural Networks created in this work predicted Openness,

Conscientiousness, and Agreeableness traits better than all the previous studies. For the Extraversion trait, previous work reported a higher accuracy of 89.1%, but this study was conducted with a population size of 4 [101]. Next best study for Extraversion trait achieved an accuracy of 76.8% [124]. In case of Neuroticism trait, previous work performed slightly better than this work with an accuracy of 80.2% [124].

While the aggregate performance between the baseline and neural network were almost indistinguishable, an analysis of the residuals indicated that the neural network was providing a noisy model of the underlying phenomenon, while the baseline was subject to substantial systemic (or non-random) error. The neural network may occasionally mis-classify aspects of a particular participant's personality, but the baseline model assumes that everyone is the same. From a practical perspective the neural network model could be useful in error tolerant applications, and as a starting point for future research. The baseline model, with its assumption of population homogeneity, is essentially useless for differentiation based on personality. However, it is worth reiterating that this type of analysis, where performance is compared against a reasonable baseline, is critical when considering non-linear models of complex behavior because linear statistical metrics can provide misleading results.

Statistical metrics such as error and accuracy cannot be used alone to assess or compare the quality of machine learning models, this was very well demonstrated using random forests and SVM models. RMSE values of the neural networks and random forests models indicated that random forests and baseline models outperformed the neural network models, but this was due to population skew present in the data. This skew likely fed into the training of the models and random forests were unable to capture the relation between extracted features and personality traits resulting in the poor generalization of data, and therefore biasing the random forests towards the norm of the dataset. This resulted in lower error values despite poor generalization. Further analysis of machine learning models using fit line plots and residual plots illustrated that neural network models were more capable of generalizing the data than random forests or baseline models.

Similarly, SVM models were created to establish the guidelines while using classification algorithms to model the personality. Results indicated that, in addition to simple statistical metrics, various analyses such as confusion matrices and ROC curves should be performed to justify the quality of classification models. This was absent in the existing literature, and statistical metrics were used to report the quality of models. Accuracies of SVM models were low ranging from 55-61%, largely due to heterogeneous data which makes it difficult for SVM to find the linear separable support vectors. A greater number of distinct features and low correlations between the dependent and independent features increase the troubles for SVM models in finding the appropriate hyperplanes.

Overall, looking at the confusion matrices for classification models and fit/residual for regression models, neural network models were more capable of generalizing the data than SVM models. The major reasons for poor generalizability of SVM models were: first, amount of effort spent on tuning the SVM models was very low compared to neural network models, second, poor selection of kernels for SVM models which

predetermines the input data structure.

This work makes three important contributions to the literature and smartphone sensing research:

Spatial Descriptors: Spatial descriptors (anchor location arrival and departure times) provide a substantial improvement to prediction accuracy compared to prior work. Authors following this work will need to consider spatial descriptors and exceed the performance.

Continuous Personality: Prior models have focused on binning each personality trait into two or three categories, then treating the problem as classification [2, 24, 30, 63, 101, 125]. This is the first research to provide a continuous representation of Big Five personality traits by creating a regression model using readily available smartphone sensor data.

Improved Methodology: Most prior works do not provide a detailed analysis of performance, or comparison to a baseline algorithm [24, 30, 90, 101, 124]. In this work, a baseline algorithm 'assume population mean' was employed and compared for performance using both aggregate measures and residuals. This work sets the standard by which personality estimation from mobile devices must be reported.

The following observations are made in this study: Conscientious people plan and do things in an organized and timely manner [48]; in this study, such participants were found exhibiting punctuality and periodicity in arriving at and departing the university, as well as consistently charging their phones. This indicates that participants had less variability in arrival, departure, and battery charging times during the study period. Time spent on campus by conscientious people was consistent, which could be indicative of well-planned behavior. Active smartphone usage for conscientiousness people was low indicating that studious participants are less likely to use smartphone continuously for longer periods of time. Surprisingly, conscientious people used scheduling applications rarely, compared to other studies [24]. This could be because conscientious people might have become experts in operating Scheduling apps and Reminder apps due to which the data collection missed logging those applications below 5 minutes duty cycle. For Openness, participants visited new places quite often, as they are interested in new things and enjoy meeting new people [48], it was also noticed that open people visited new routers more often. Open-minded people used the camera and maps more, probably because of their interest in visiting and recording new places. Extroverts, characterized by talkative and outgoing attitudes, were found to use social media, made more calls, and used SMS more than others in our study. This concurred with previous studies [24, 25, 108] and our expectations. Among the five traits, the Neuroticism trait was difficult to model since the participants were university students whose daily routines and activities were predetermined by the school work and therefore, emotional stability levels might have been less apparent in daily activities.

This work leveraged an important dataset, which had spatial and communication behaviors for 1461 days from 80 people, to establish its findings. This is sufficient for the proof of concept and demonstration of feasibility which is the primary contribution of this thesis; however, the size and structure of the participants' personality profiles do not fully reflect the general population. In particular, the population was skewed

toward higher conscientiousness and agreeableness, which is not unexpected for a university population. Additionally, there was a significant class skew towards the mean of the sample, with poor representation at the extremes of any of the personality dimensions. This skew inflated the efficacy of the baseline algorithm, as the greater density of average participants reduced the error contribution of the outliers. Because the population under consideration was from the university community, everyone had similar constraints on their schedules, characterized by less rigid arrival and departure times than many professions, and classes scheduled primarily during typical North American working hours. For these reasons the generalizability of the generated models are limited, but the generalizability of the approach should hold for larger and more diverse populations.

There are several limitations to the work that point towards future research. First, as noted in the preceding paragraph, the generalizability of findings is limited by the demographic, both in terms of the personalities represented and the behaviors exhibited. Further studies with broader demographic cross-sections should be undertaken before the validity of the approach can be fully realized. Second, data collection duty cycle of 5 minutes might have missed logging usage of applications used for less than 5 minutes per session. Third, the machine learning algorithm (neural networks) employed, while effective at creating models, has poor interpretability, acting as a black box. While limited success was noticed with statistical regression and Bayesian models before attempting the neural network, more sophisticated statistical models might provide similar precision and greater confidence in mapping causal pathways. Absent sufficient statistical models, sensitivity analysis could be performed on the existing model to determine the impact of including various features for estimating particular personality traits. Finally, this work only scratched the surface of potential geo-spatial features which may correlate with personality. Additional features, such as the amount of time spent in different types of shopping establishments, the number of hours spent at home, at work, or elsewhere, and even the overall predictability of an individual's trajectory may provide potentially more discriminant features for distinguishing personality from behavior. Further studies investigating other features are warranted.

7.1 Summary

This work demonstrated that it is possible, over a period of several days, to estimate the personality inventory of a participant using behavioral data sensed from commodity smartphones. This work offers a new framework by using both smartphone usage and also location-aware techniques to find the personality without the need of parsing the app-specific content and social media content. Incorporating aspects of spatial behavior improved performance over what has previously been reported in the literature. This research has potential applications in the social sciences and public health, as a way of automatically identifying a potential confound, in marketing as a way of automatically extracting broader behaviors, and in location-based services as a method of extracting preferences for personalized experiences. While promising, this work represents an initial step

and proof of concept, and substantial research work remains in extending, validating and applying this research across a number of fields.

References

[1] Yair Amichai-Hamburger and Gideon Vinitzky. Social network use and personality. *Computers in Human Behavior*, 26(6):1289–1295, 2010.

[2] Margit Antal, László Zsolt Szabó, and Győző Nemes. Predicting user identity and personality traits from mobile sensor data. In *International Conference on Information and Software Technologies*, pages 163–173. Springer, 2016.

[3] Siamak Aram, Amedeo Troiano, and Eros Pasero. Environment sensing using smartphone. In *Sensors Applications Symposium (SAS), 2012 IEEE*, pages 1–4. IEEE, 2012.

[4] Yuki Arase, Fei Ren, and Xing Xie. User activity understanding from mobile phone sensors. In *Proceedings of the 12th ACM International Conference Adjunct Papers on Ubiquitous Computing-Adjunct*, pages 391–392. ACM, 2010.

[5] Daniel Ashbrook and Thad Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5):275–286, 2003.

[6] Elizabeth J Austin, Donald H Saklofske, and Vincent Egan. Personality, well-being and health correlates of trait emotional intelligence. *Personality and Individual Differences*, 38(3):547–558, 2005.

[7] S Santhosh Baboo and I Kadar Shereef. An efficient weather forecasting system using artificial neural network. *International Journal of Environmental Science and Development*, 1(4):321, 2010.

[8] Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell. Personality and patterns of facebook usage. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 24–32. ACM, 2012.

[9] Mitja D Back, Juliane M Stopfer, Simine Vazire, Sam Gaddis, Stefan C Schmukle, Boris Egloff, and Samuel D Gosling. Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21(3):372–374, 2010.

[10] Murray R Barrick and Michael K Mount. The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, 44(1):1–26, 1991.

[11] Murray R Barrick, Greg L Stewart, Mitchell J Neubert, and Michael K Mount. Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, 83(3):377, 1998.

[12] Imad A Basheer and M Hajmeer. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43(1):3–31, 2000.

[13] Roy F Baumeister, Kathleen D Vohs, and David C Funder. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2(4):396–403, 2007.

[14] Dror Ben-Zeev, Christopher J Brenner, Mark Begale, Jennifer Duffecy, David C Mohr, and Kim T Mueser. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. *Schizophrenia Bulletin*, 40(6):1244–1253, 2014.

[15] Joan-Isaac Biel, Oya Aran, and Daniel Gatica-Perez. You are known by how you vlog: Personality impressions and nonverbal behavior in youtube. In *ICWSM*, 2011.

[16] Bjorn. Personality and obesity. `http://www.bjornboman.com/2016/01/16/personality-and-obesity/`, 2016. [Online; accessed Aug 2018].

[17] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[18] Tom Buchanan and John L Smith. Using the internet for psychological research: Personality testing on the world wide web. *British Journal of Psychology*, 90(1):125–144, 1999.

[19] Michelle Nicole Burns, Mark Begale, Jennifer Duffecy, Darren Gergle, Chris J Karr, Emily Giangrande, and David C Mohr. Harnessing context sensing to develop a mobile intervention for depression. *Journal of Medical Internet Research*, 13(3), 2011.

[20] Sarah Butt and James G Phillips. Personality and self reported mobile phone use. *Computers in Human Behavior*, 24(2):346–360, 2008.

[21] Raymond Bernard Cattell, Herbert W Eber, and Maurice M Tatsuoka. *Handbook for the sixteen personality factor questionnaire (16 PF): In clinical, educational, industrial, and research psychology, for use with all forms of the test.* Institute for Personality and Ability Testing, 1970.

[22] Tianfeng Chai and Roland R Draxler. Root mean square error (rmse) or mean absolute error (mae)?– arguments against avoiding rmse in the literature. *Geoscientific Model Development*, 7(3):1247–1250, 2014.

[23] Olivier Chapelle, Patrick Haffner, and Vladimir N Vapnik. Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, 10(5):1055–1064, 1999.

[24] Gokul Chittaranjan, Jan Blom, and Daniel Gatica-Perez. Who's who with big-five: Analyzing and classifying personality traits with smartphones. In *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*, pages 29–36. IEEE, 2011.

[25] Teresa Correa, Amber Willard Hinsley, and Homero Gil De Zuniga. Who interacts on the web?: The intersection of users personality and social media use. *Computers in Human Behavior*, 26(2):247–253, 2010.

[26] Paul T Costa and Robert R MacCrae. *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual.* Psychological Assessment Resources, Incorporated, 1992.

[27] Paul T Costa and Robert R McCrae. The neo personality inventory. 1985.

[28] William Grant Dahlstrom and George Schlager Welsh. An mmpi handbook: A guide to use in clinical practice and research. 1960.

[29] Robert N Davis. Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments, & Computers*, 31(4):572–577, 1999.

[30] Yves-Alexandre de Montjoye, Jordi Quoidbach, Florent Robic, and Alex Pentland. Predicting personality using novel mobile phone-based metrics. In *SBP*, pages 48–55. Springer, 2013.

[31] Rodrigo de Oliveira, Alexandros Karatzoglou, Pedro Concejero Cerezo, Ana Armenta Lopez de Vicuña, and Nuria Oliver. Towards a psychographic user model from mobile phone usage. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 2191–2196. ACM, 2011.

[32] John M Digman. Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41(1):417–440, 1990.

[33] Trinh Minh Tri Do and Daniel Gatica-Perez. The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. *IEEE Transactions on Mobile Computing*, 13(3):638–648, 2014.

[34] Harris Drucker, Donghui Wu, and Vladimir N Vapnik. Support vector machines for spam categorization. *IEEE Transactions on Neural Networks*, 10(5):1048–1054, 1999.

[35] Moustafa Elhamshary and Moustafa Youssef. Semsense: Automatic construction of semantic indoor floorplans. In *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*, pages 1–11. IEEE, 2015.

[36] Nicole B Ellison, Charles Steinfield, and Cliff Lampe. The benefits of facebook friends: social capital and college students use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, 2007.

[37] Elisabeth Engelberg and Lennart Sjöberg. Internet use, social skills, and adjustment. *Cyber Psychology & Behavior*, 7(1):41–47, 2004.

[38] Inc Ethica Data Services. Ethica Data Services. `www.ethicadata.com`, 2010. [Online; accessed Jan 2018].

[39] Hans Jurgen Eysenck and SGB Eysenck. The eysenck personality inventory. 1965.

[40] Katayoun Farrahi and Daniel Gatica-Perez. What did you do today?: discovering daily routines from large-scale mobile data. In *Proceedings of the 16th ACM International Conference on Multimedia*, pages 849–852. ACM, 2008.

[41] Daniel Fuller, Martine Shareck, and Kevin Stanley. Ethical implications of location and accelerometer measurement in health research studies with mobile sensing devices. *Social Science & Medicine*, 191:84–88, 2017.

[42] R Michael Furr. Personality psychology as a truly behavioural science. *European Journal of Personality*, 23(5):369–401, 2009.

[43] Aurélien Géron. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems.* " O'Reilly Media, Inc.", 2017.

[44] ATC Goh. Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 9(3):143–151, 1995.

[45] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting personality from twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 149–156. IEEE, 2011.

[46] Jennifer Golbeck, Cristina Robles, and Karen Turner. Predicting personality with social media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 253–262. ACM, 2011.

[47] Lewis R Goldberg. An alternative" description of personality": the big-five factor structure. *Journal of Personality and Social Psychology*, 59(6):1216, 1990.

[48] Lewis R Goldberg. The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1):26, 1992.

[49] Lewis R Goldberg and Gerard Saucier. What is beyond the big five? *Journal of Personality*, 66(4):495–524, 1998.

[50] Samuel D Gosling, Oliver P John, Kenneth H Craik, and Richard W Robins. Do people know how they behave? self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, 74(5):1337, 1998.

[51] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *Acoustics, Speech and Signal Processing (icassp), 2013 IEEE International Conference on*, pages 6645–6649. IEEE, 2013.

[52] Agnes Grünerbl, Amir Muaremi, Venet Osmani, Gernot Bahle, Stefan Oehler, Gerhard Tröster, Oscar Mayora, Christian Haring, and Paul Lukowicz. Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE Journal of Biomedical and Health Informatics*, 19(1):140–148, 2015.

[53] Rosanna E Guadagno, Bradley M Okdie, and Cassie A Eno. Who blogs? personality predictors of blogging. *Computers in Human Behavior*, 24(5):1993–2004, 2008.

[54] Kevin Gurney. *An introduction to neural networks*. CRC press, 2014.

[55] Yair Amichai Hamburger and Elisheva Ben-Artzi. The relationship between extraversion and neuroticism and the different uses of the internet. *Computers in Human Behavior*, 16(4):441–449, 2000.

[56] Gabriella M Harari, Nicholas D Lane, Rui Wang, Benjamin S Crosier, Andrew T Campbell, and Samuel D Gosling. Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, 11(6):838–854, 2016.

[57] Mohammad S Hashemian, Kevin G Stanley, Dylan L Knowles, Jonathan Calver, and Nathaniel D Osgood. Human network data collection in the wild: the epidemiological utility of micro-contact and location data. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, pages 255–264. ACM, 2012.

[58] Peter Hills and Michael Argyle. Uses of the internet and their relationships with individual differences in personality. *Computers in Human Behavior*, 19(1):59–70, 2003.

[59] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[60] Gregory M Hurtz and John J Donovan. Personality and job performance: The big five revisited. *Journal of Applied Psychology*, 85(6):869–879, 2000.

[61] Ozlem Durmaz Incel, Mustafa Kose, and Cem Ersoy. A review and taxonomy of activity recognition on mobile phones. *BioNanoScience*, 3(2):145–171, 2013.

[62] Anil K Jain, Jianchang Mao, and K Moidin Mohiuddin. Artificial neural networks: A tutorial. *Computer*, 29(3):31–44, 1996.

[63] Natasha Jaques, Sara Taylor, Asaph Azaria, Asma Ghandeharioun, Akane Sano, and Rosalind Picard. Predicting students' happiness from physiology, phone, mobility, and behavioral data. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*, pages 222–228. IEEE, 2015.

[64] Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *European Conference on Machine Learning*, pages 137–142. Springer, 1998.

[65] Oliver P John, Eileen M Donahue, and Robert L Kentle. The big five inventory versions 4a and 54, 1991.

[66] Oliver P John, Laura P Naumann, and Christopher J Soto. Paradigm shift to the integrative big five trait taxonomy. *Handbook of Personality: Theory and Research*, 3(2):114–158, 2008.

[67] Timothy A Judge, Chad A Higgins, Carl J Thoresen, and Murray R Barrick. The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52(3):621–652, 1999.

[68] Andreas M Kaplan and Michael Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, 53(1):59–68, 2010.

[69] Yakup Kara, Melek Acar Boyacioglu, and Ömer Kaan Baykan. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. *Expert Systems with Applications*, 38(5):5311–5319, 2011.

[70] Arvid Karsvall. Personality preferences in graphical interface design. In *Proceedings of the Second Nordic Conference on Human-Computer Interaction*, pages 217–218. ACM, 2002.

[71] Susan L Kichuk and Willi H Wiesner. The big five personality factors and team performance: implications for selecting successful product design teams. *Journal of Engineering and Technology Management*, 14(3-4):195–221, 1997.

[72] Takashi Kimoto, Kazuo Asakawa, Morio Yoda, and Masakazu Takeoka. Stock market prediction system with modular neural networks. In *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*, pages 1–6. IEEE, 1990.

[73] Robert Kraut, Michael Patterson, Vicki Lundmark, Sara Kiesler, Tridas Mukophadhyay, and William Scherlis. Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53(9):1017, 1998.

[74] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[75] Richard N Landers and John W Lounsbury. An investigation of big five and narrow personality traits in relation to internet usage. *Computers in Human Behavior*, 22(2):283–293, 2006.

[76] Kwan Min Lee and Clifford Nass. Designing social presence of social actors in human computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 289–296. ACM, 2003.

[77] Lin Liao, Dieter Fox, and Henry Kautz. Location-based activity recognition. In *Advances in Neural Information Processing Systems*, pages 787–794, 2006.

[78] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, pages 389–402. ACM, 2013.

[79] Armand W Loranger, Norman Sartorius, Antonio Andreoli, Peter Berger, Peter Buchheim, SM Channabasavanna, Bina Coid, Alv Dahl, Rene FW Diekstra, Brian Ferguson, et al. The international personality disorder examination: The world health organization/alcohol, drug abuse, and mental health administration international pilot study of personality disorders. *Archives of General Psychiatry*, 51(3):215–224, 1994.

[80] David D Luxton, Russell A McCann, Nigel E Bush, Matthew C Mishkind, and Greg M Reger. mhealth for mental health: Integrating smartphone technology in behavioral healthcare. *Professional Psychology: Research and Practice*, 42(6):505, 2011.

[81] Rischan Mafrur, I Gde Dharma Nugraha, and Deokjai Choi. Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose. *Human-centric Computing and Information Sciences*, 5(1):31, 2015.

[82] François Mairesse, Marilyn A Walker, Matthias R Mehl, and Roger K Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30:457–500, 2007.

[83] Gloria Mark and Yoav Ganzach. Personality and internet usage: A large-scale representative study of young adults. *Computers in Human Behavior*, 36:274–281, 2014.

[84] Carol McDonald. Predicting Loan Credit Risk using Apache Spark Machine Learning Random Forests. `https://mapr.com/blog/predicting-loan-credit-risk-using-apache-spark-machine-learning-random-forests/`, 2016. [Online; accessed Aug 2018].

[85] James C McElroy, Anthony R Hendrickson, Anthony M Townsend, and Samuel M DeMarie. Dispositional factors in internet use: personality versus cognitive style. *MIS Quarterly*, pages 809–820, 2007.

[86] Evangelia Micheli-Tzanakou, Hassan Sheikh, and Bin Zhu. Neural networks and blood cell identification. *Journal of Medical Systems*, 21(4):201–210, 1997.

[87] Emiliano Miluzzo, Nicholas D Lane, Kristóf Fodor, Ronald Peterson, Hong Lu, Mirco Musolesi, Shane B Eisenman, Xiao Zheng, and Andrew T Campbell. Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, pages 337–350. ACM, 2008.

[88] Mitek. 2014 Zogby Analytics - Millennial Study. `https://www.miteksystems.com/resources/2014-zogby-analytics-millennial-study`, 2014. [Online; accessed Jan 2018].

[89] Prashanth Mohan, Venkata N Padmanabhan, and Ramachandran Ramjee. Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, pages 323–336. ACM, 2008.

[90] Bjarke Mønsted, Anders Mollgaard, and Joachim Mathiesen. Phone-based metric as a predictor for basic personality traits. *ArXiv Preprint ArXiv:1604.04696*, 2016.

[91] Raul Montoliu, Jan Blom, and Daniel Gatica-Perez. Discovering places of interest in everyday life from smartphone data. *Multimedia Tools and Applications*, 62(1):179–207, 2013.

[92] Min Mun, Sasank Reddy, Katie Shilton, Nathan Yau, Jeff Burke, Deborah Estrin, Mark Hansen, Eric Howard, Ruth West, and Péter Boda. Peir, the personal environmental impact report, as a platform for participatory sensing systems research. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, pages 55–68. ACM, 2009.

[93] Isabel Briggs Myers. The myers-briggs type indicator: Manual (1962). 1962.

[94] Maria Augusta SN Nunes, Stefano A Cerri, and Nathalie Blanc. Towards user psychological profile. In *Proceedings of the VIII Brazilian symposium on human factors in computing systems*, pages 196–203. Sociedade Brasileira de Computação, 2008.

[95] Rita Orji, Lennart E Nacke, and Chrysanne Di Marco. Towards personality-driven persuasive health games and gamified systems. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1015–1027. ACM, 2017.

[96] Melissa C OConnor and Sampo V Paunonen. Big five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, 43(5):971–990, 2007.

[97] Karen A Pasveer and John H Ellard. The making of a personality inventory: Help from the www. *Behavior Research Methods, Instruments, & Computers*, 30(2):309–313, 1998.

[98] Delroy L Paulhus and Simine Vazire. The self-report method. *Handbook of Research Methods in Personality Psychology*, 1:224–239, 2007.

[99] Sampo V Paunonen and Douglas N Jackson. What is beyond the big five? plenty! *Journal of Personality*, 68(5):821–835, 2000.

[100] James G Phillips, Sarah Butt, and Alex Blaszczynski. Personality and self-reported use of mobile phones for games. *Cyber Psychology & Behavior*, 9(6):753–758, 2006.

[101] Fabio Pianesi, Nadia Mana, Alessandro Cappelletti, Bruno Lepri, and Massimo Zancanaro. Multi-modal recognition of personality traits in social interactions. In *Proceedings of the 10th International Conference on Multimodal Interfaces*, pages 53–60. ACM, 2008.

[102] Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. Our twitter profiles, our selves: Predicting personality with twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 180–185. IEEE, 2011.

[103] Rajib Kumar Rana, Chun Tung Chou, Salil S Kanhere, Nirupama Bulusu, and Wen Hu. Ear-phone: an end-to-end participatory urban noise mapping system. In *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pages 105–116. ACM, 2010.

[104] Andreas Rauch and Michael Frese. Let's put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners' personality traits, business creation, and success. *European Journal of Work and Organizational Psychology*, 16(4):353–385, 2007.

[105] Apostolos Nicholas Refenes, Achileas Zapranis, and Gavin Francis. Stock performance modeling using neural networks: a comparative study with regression models. *Neural Networks*, 7(2):375–388, 1994.

[106] John P Robinson, Phillip R Shaver, and Lawrence S Wrightsman. *Measures of Personality and Social Psychological Attitudes: Measures of Social Psychological Attitudes*, volume 1. Academic Press, 2013.

[107] Craig Ross, Emily S Orr, Mia Sisic, Jaime M Arseneault, Mary G Simmering, and R Robert Orr. Personality and motivations associated with facebook use. *Computers in Human Behavior*, 25(2):578–586, 2009.

[108] Tracii Ryan and Sophia Xenos. Who uses facebook? an investigation into the relationship between the big five, shyness, narcissism, loneliness, and facebook usage. *Computers in Human Behavior*, 27(5):1658–1664, 2011.

[109] Batul Saati, May Salem, and Willem-Paul Brinkman. Towards customized user interface skins: investigating user personality and skin colour. *Proceedings of HCI 2005*, 2:89–93, 2005.

[110] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of Medical Internet Research*, 17(7), 2015.

[111] Jesus F Salgado. The five factor model of personality and job performance in the european community. *Journal of Applied Psychology*, 82(1):30, 1997.

[112] Sandhya Samarasinghe. *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*. Auerbach Publications, 2016.

[113] Suranga Seneviratne, Aruna Seneviratne, Prasant Mohapatra, and Anirban Mahanti. Predicting user traits from a snapshot of apps installed on a smartphone. *ACM SIGMOBILE Mobile Computing and Communications Review*, 18(2):1–8, 2014.

[114] Sandra Servia-Rodríguez, Kiran K Rachuri, Cecilia Mascolo, Peter J Rentfrow, Neal Lathia, and Gillian M Sandstrom. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In *Proceedings of the 26th International Conference on World Wide Web*, pages 103–112. International World Wide Web Conferences Steering Committee, 2017.

[115] Jianqiang Shen, Oliver Brdiczka, and Juan Liu. Understanding email writers: Personality prediction from email messages. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 318–330. Springer, 2013.

[116] Statista. Number of smartphone users worldwide from 2014 to 2020 (in billions). `https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/`, 2016. [Online; accessed Jan 2018].

[117] Ingo Steinwart and Andreas Christmann. *Support vector machines*. Springer Science & Business Media, 2008.

[118] Stanley Sue and Derald W Sue. Chinese-american personality and mental health. *Amerasia Journal*, 1(2):36–49, 1971.

[119] Antonio Terracciano and Paul T Costa. Smoking and the five-factor model of personality. *Addiction*, 99(4):472–481, 2004.

[120] Arvind Thiagarajan, Lenin Ravindranath, Katrina LaCurts, Samuel Madden, Hari Balakrishnan, Sivan Toledo, and Jakob Eriksson. Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 85–98. ACM, 2009.

[121] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 3–14. ACM, 2014.

[122] Ting Wei and Scott Bell. Impact of indoor location information reliability on users trust of an indoor positioning system. *Geographic Information Science*, pages 258–269, 2012.

[123] Uwe Wolfradt and Jean E Pretz. Individual differences in creativity: Personality, story writing, and hobbies. *European Journal of Personality*, 15(4):297–310, 2001.

[124] William R Wright and David N Chin. Personality profiling from text: introducing part-of-speech n-grams. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 243–253. Springer, 2014.

[125] Runhua Xu, Remo Manuel Frey, Elgar Fleisch, and Alexander Ilic. Understanding the impact of personality traits on mobile app adoption–insights from a large-scale field study. *Computers in Human Behavior*, 62:244–256, 2016.

[126] Wu Youyou, Michal Kosinski, and David Stillwell. Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4):1036–1040, 2015.

[127] Hao Zhao and Scott E Seibert. The big five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology*, 91(2):259, 2006.

Appendix A

Neural Networks

A.1 Models

Neural network models for Conscientiousness, Extraversion, Agreeableness, and Neuroticism traits are presented in Figure A.1, and Figure A.2.
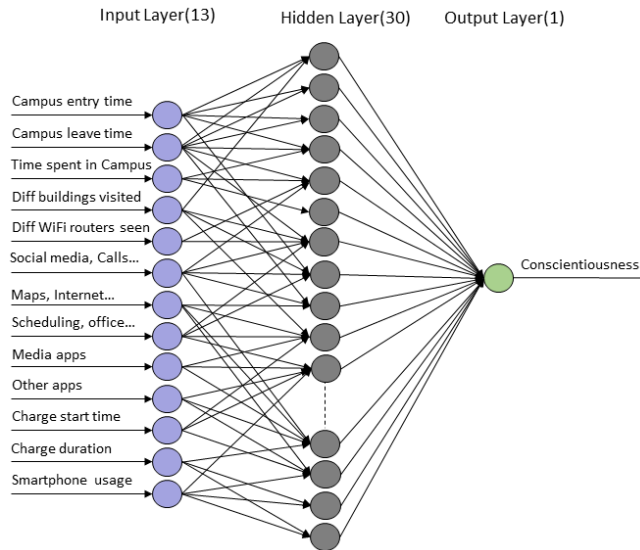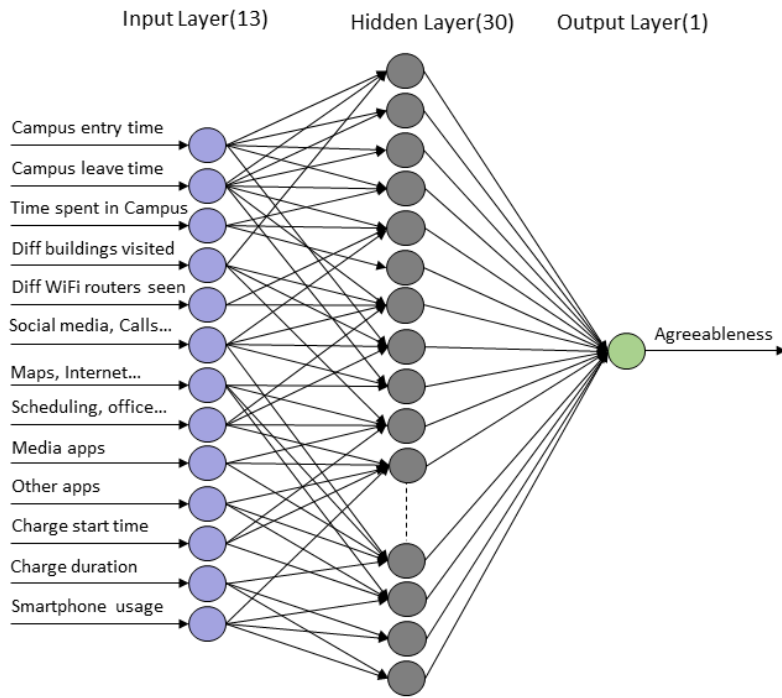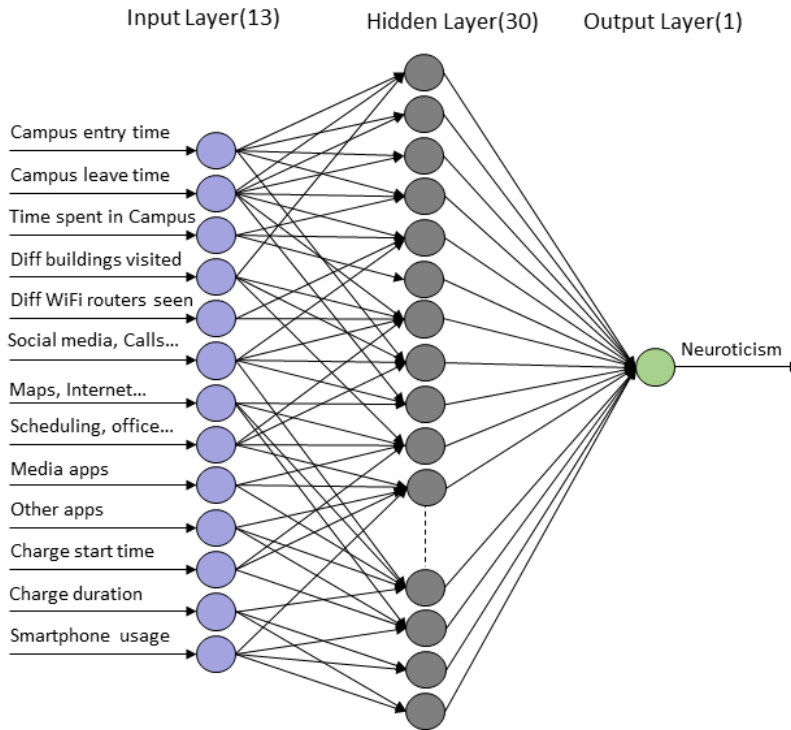

(a) Conscientiousness
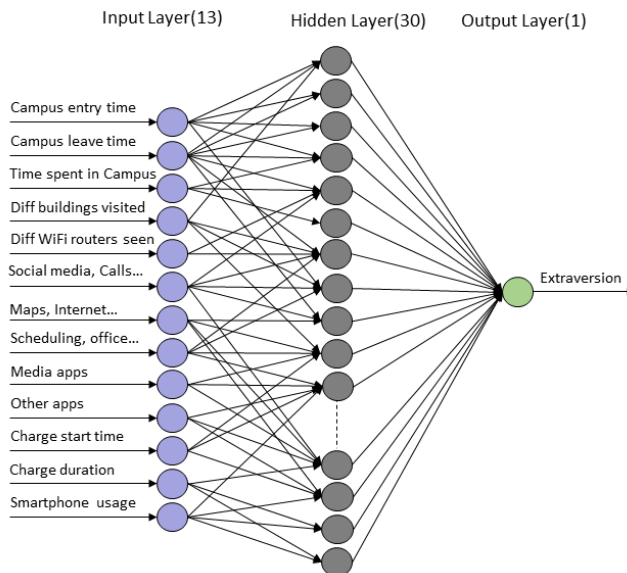
(b) Extraversion

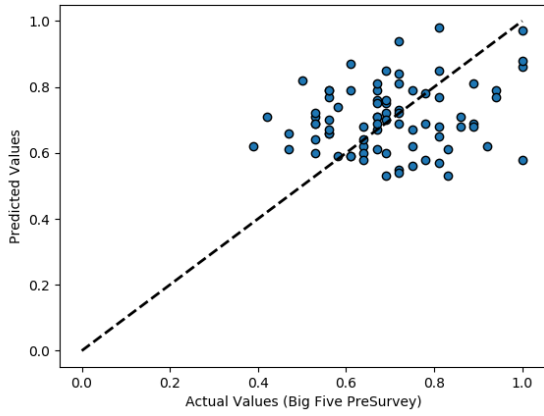**Figure A.1:** Neural Network models for Conscientiousness and Extraversion

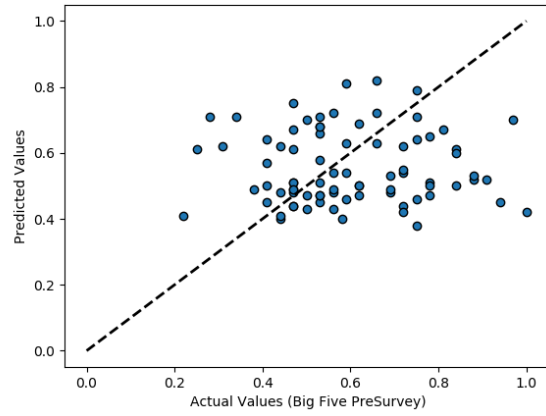**(a)** Agreeableness

(b) Neuroticism

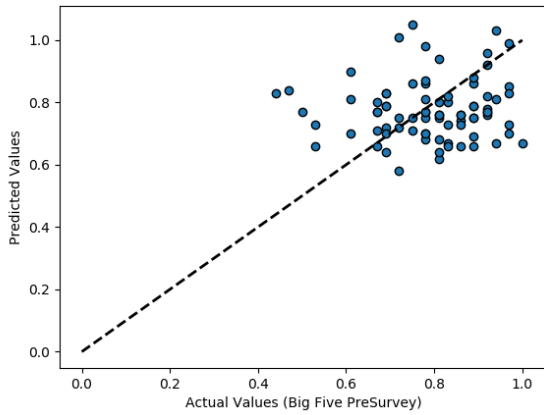**Figure A.2:** Neural Network models for Agreeableness and Neuroticism

A.2 Training and Validation

Fit line plots of neural network models for activation functions ReLU, Tanh and Sigmoid are presented in Figure A.3, Figure A.4 and Figure A.5 respectively.

(a) Conscientiousness

(b) Extraversion
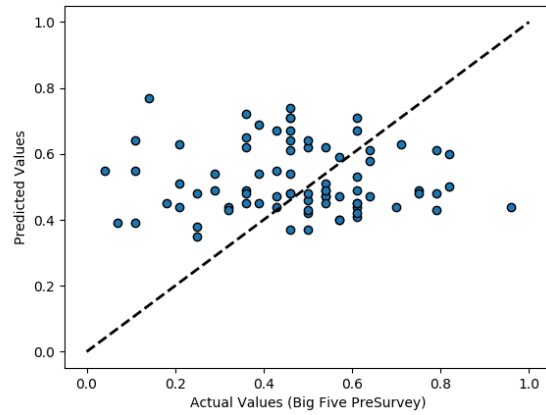
(c) Agreeableness

(d) Neuroticism

Figure A.3: Fit Line plots of Neural Networks for ReLU activation Function
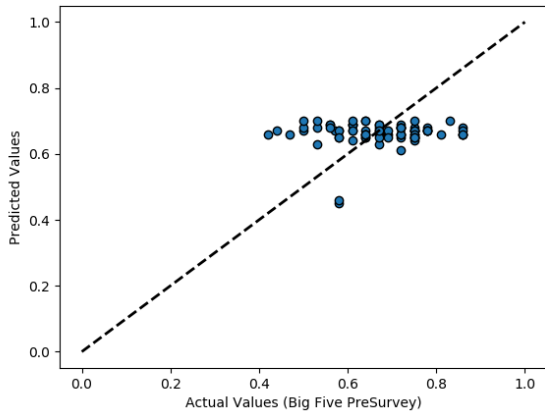
(a) Openness
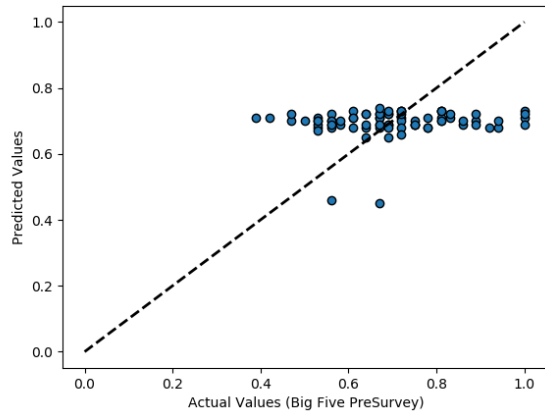
(b) Conscientiousness
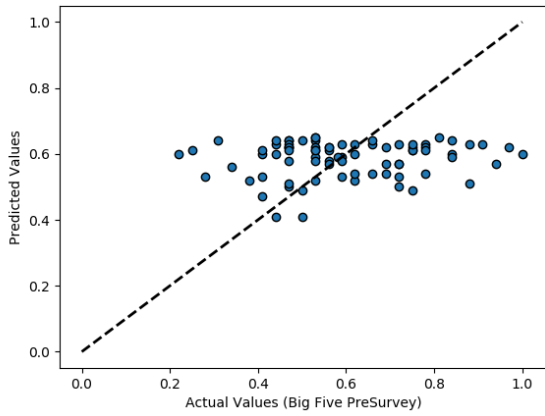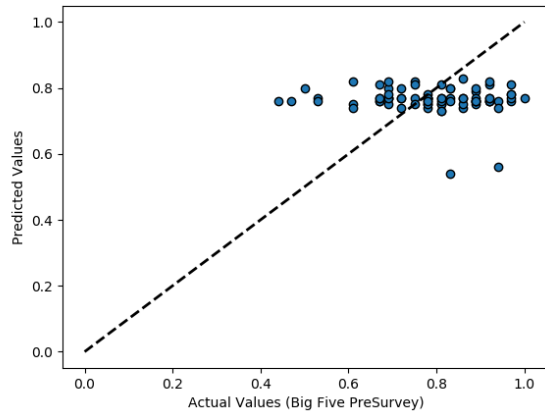
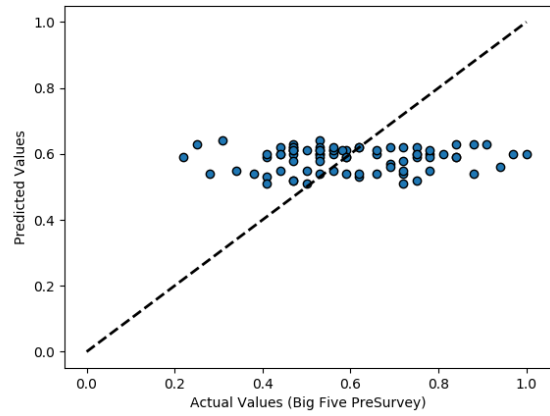**(c)** Extraversion
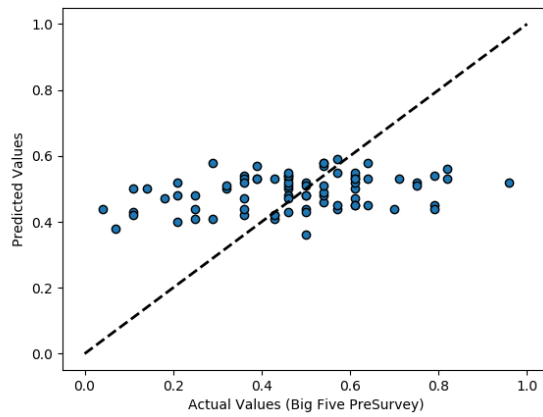
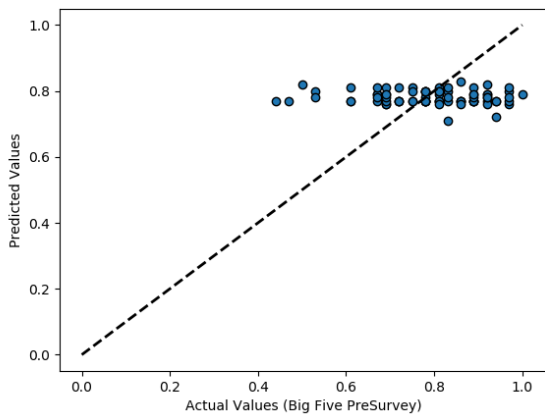**(d)** Agreeableness

(e) Neuroticism

Figure A.4: Fit Line plots of Neural Networks for Tanh activation Function
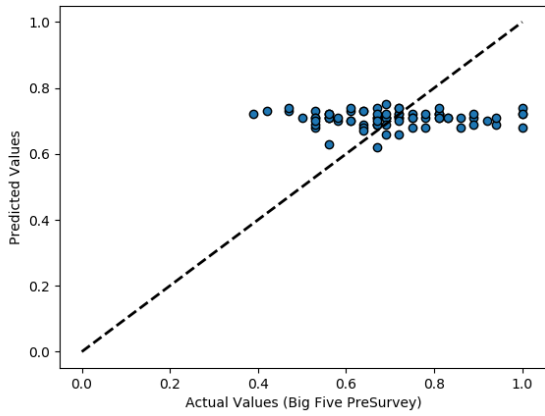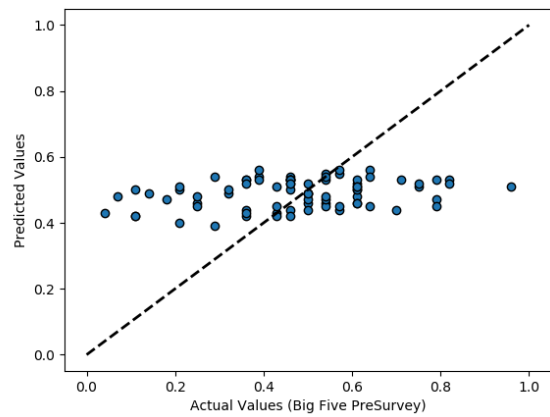
(a) Conscientiousness

(b) Extraversion

(c) Agreeableness

(d) Neuroticism

Figure A.5: Fit Line plots of Neural Networks for Sigmoid activation Function