

ABSTRACT

This study centers on personality prediction. The purpose of this study is to develop an Artificial Neural Network model that can be used to predict a person's big five personality based on only their Facebook activity. Everyday social media for instance Facebook, experiences a rapid increase in usage and popularity. Various people see social media e.g. Facebook as a medium to share and obtain a variety of information and also as a platform to stay updated. Facebook today provides loads of information concerning user's daily interactions. Various researchers and studies harness the streams of information on these social media platforms as an important asset to better understand human behavior, social interaction and personality. Numerous researches have been conducted in this field and even now it continues to grow. These studies have been able to use these studies to better understand who the users are, understand what their interest is and what they need. Information as these is important to businesses to better understand their clients. Also, law enforcement agencies can predict potential threats to the society with this information. The aim of this study is to build a predictive model that uses Facebook user's data and activity to predict the big 5 personalities. In order to do this, this study combines the inference features highlighted in three different studies which are the number of likes, events, groups, tags, updates, network size, relationship status, age and gender. The study was conducted on 7438 unique Facebook participants gotten from the myPersonality database. The findings of this study showed how much a person's personality can be predicted only by analyzing their Facebook activity. The ANN model was able to correctly classify an individual's personality at an 85% prediction accuracy. This study proposes a model by combining inference features from three different studies and predicts personality based on these features alone without including words or contents of status updates differing it from other studies.

Keywords: Artificial Neural Network; Big 5 personality; Facebook; Machine Learning; Personality Prediction

ÖZET

Bu çalışma kişilik tahmini üzerinde merkezi. Bu çalışmanın amacı, bir kişinin “büyük 5” kişiliğini sadece Facebook etkinliklerine dayanarak tahmin etmek için kullanılabilecek Yapay Sinir Ağı modelini geliştirmektir. Çeşitli insanlar sosyal medya örneğin Facebook paylaşmak ve bilgi çeşitli elde etmek için bir orta olarak görmek ve onlar da bir platform olarak güncel kalmak için görmek. Bu günlerde, Facebook bir kullanıcının günlük etkileşimleri hakkında birçok bilgi sunuyor. Çeşitli araştırmacılar ve çalışmalar insan davranışlarını, sosyal etkileşimi ve kişiliği daha iyi anlamak için önemli bir varlık olarak bu sosyal medya platformları hakkında bol bilgi kullanır. Bu alanda çok sayıda araştırma yapıldı ve şimdi bile büyümeye devam ediyor. Bu çalışmalar daha iyi kullanıcıların kim olduğunu anlamak için kendi çalışmalarını kullanmak başardık, ne onların ilgi ve ne ihtiyaç duydukları anlamak. Bunlar gibi bilgiler, işletmelerin müşterilerine daha iyi anlaşılması için önemlidir. Kanun uygulama kurumları bu bilgilerle topluma potansiyel tehditleri tahmin edebilir. Bu çalışmanın amacı, büyük 5 kişiliği tahmin etmek için Facebook Kullanıcı veri ve aktivite kullanan bir öngörü modeli inşa etmektir. Bunu yapmak için, bu çalışma, beğeni, etkinlik, grup, etiket, güncelleme, ağ boyutu, ilişki durumu, yaş ve cinsiyet sayısı olan üç farklı çalışmada vurgulanan özellikleri bir araya getirmektedir. Çalışma myPersonality veritabanından alınan 7438 benzersiz Facebook katılımcısı üzerinde yapıldı. Bu çalışmanın bulguları, bir kişinin kişiliği sadece Facebook aktivitesini analiz ederek tahmin edilebilir ne kadar gösterdi. Ann modeli doğru bir 85% tahmin doğruluğu ile bireyin kişiliği sınıflandırmak başardı. Bu çalışmada üç farklı çalışmada türetilen özellikleri birleştirerek bir model öneriyor ve kelime veya durum güncellemeleri içeriği dahil olmadan tek başına bu özelliklere dayalı kişilik tahmin.

Anahtar kelimeler: Büyük 5 Kişilik; Facebook; Kişilik tahmini; Makine öğrenme; Yapay sinir ağı;

**PREDICTING PERSONALITY FROM FACEBOOK
DATA: A NEURAL NETWORK APPROACH**

**A THESIS SUBMITTED TO THE GRADUATE
SCHOOL OF APPLIED SCIENCES
OF
NEAR EAST UNIVERSITY**

**By
OBINNA HARRISON EJIMOGU**

**In Partial Fulfillment of the Requirements for
The Degree of Master of Science
in
Computer Information Systems**

NICOSIA, 2018

OBINNA H. EJIMOGU

**PREDICTING PERSONALITY FROM FACEBOOK DATA: A NEURAL
NETWORK APPROACH**

**NEU
2018**

**PREDICTING PERSONALITY FROM FACEBOOK
DATA: A NEURAL NETWORK APPROACH**

**A THESIS SUBMITTED TO THE GRADUATE
SCHOOL OF APPLIED SCIENCES
OF
NEAR EAST UNIVERSITY**

**By
OBINNA H. EJIMOGU**

**In Partial Fulfillment of the Requirements for
the Degree of Master of Science
in
Computer Information Systems**

NICOSIA, 2018

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name:

Signature:

Date:

To my family...

ACKNOWLEDGMENTS

First and foremost, I give a heartfelt thanks to an amazing and understanding supervisor Assist. Prof. Dr. Seren Başaran for her wonderful support, directions and for providing me with all the required skills and research tools to start and complete this study within the stipulated time. Secondly to Prof. Dr. Nadire Cavus for initial administrative guide on what it takes to complete this study. I also want to appreciate Prof. Dr. Adnan Khashman for his invaluable contribution towards this completion of this study.

Finally I appreciate my parents Mr. Ndubuisi and Mrs. Ngozi Ejimogu especially for their unwavering love and constant support and siblings who kept pushing for success and for their constant support and prayers for what I need to finish well.

ABSTRACT

This study centers on personality prediction. The purpose of this study is to develop an Artificial Neural Network model that can be used to predict a person's big five personality based on only their Facebook activity. Everyday social media for instance Facebook, experiences a rapid increase in usage and popularity. Various people see social media e.g. Facebook as a medium to share and obtain a variety of information and also as a platform to stay updated. Facebook today provides loads of information concerning user's daily interactions. Various researchers and studies harness the streams of information on these social media platforms as an important asset to better understand human behavior, social interaction and personality. Numerous researches have been conducted in this field and even now it continues to grow. These studies have been able to use these studies to better understand who the users are, understand what their interest is and what they need. Information as these is important to businesses to better understand their clients. Also, law enforcement agencies can predict potential threats to the society with this information. The aim of this study is to build a predictive model that uses Facebook user's data and activity to predict the big 5 personalities. In order to do this, this study combines the inference features highlighted in three different studies which are the number of likes, events, groups, tags, updates, network size, relationship status, age and gender. The study was conducted on 7438 unique Facebook participants gotten from the myPersonality database. The findings of this study showed how much a person's personality can be predicted only by analyzing their Facebook activity. The ANN model was able to correctly classify an individual's personality at an 85% prediction accuracy. This study proposes a model by combining inference features from three different studies and predicts personality based on these features alone without including words or contents of status updates differing it from other studies.

Keywords: Artificial Neural Network; Big 5 personality; Facebook; Machine Learning; Personality Prediction

ÖZET

Bu çalışma kişilik tahmini üzerinde merkezi. Bu çalışmanın amacı, bir kişinin “büyük 5” kişiliğini sadece Facebook etkinliklerine dayanarak tahmin etmek için kullanılabilecek Yapay Sinir Ağı modelini geliştirmektir. Çeşitli insanlar sosyal medya örneğin Facebook paylaşmak ve bilgi çeşitli elde etmek için bir orta olarak görmek ve onlar da bir platform olarak güncel kalmak için görmek. Bu günlerde, Facebook bir kullanıcının günlük etkileşimleri hakkında birçok bilgi sunuyor. Çeşitli araştırmacılar ve çalışmalar insan davranışlarını, sosyal etkileşimi ve kişiliği daha iyi anlamak için önemli bir varlık olarak bu sosyal medya platformları hakkında bol bilgi kullanır. Bu alanda çok sayıda araştırma yapıldı ve şimdi bile büyümeye devam ediyor. Bu çalışmalar daha iyi kullanıcıların kim olduğunu anlamak için kendi çalışmalarını kullanmak başardık, ne onların ilgi ve ne ihtiyaç duydukları anlamak. Bunlar gibi bilgiler, işletmelerin müşterilerine daha iyi anlaşılması için önemlidir. Kanun uygulama kurumları bu bilgilerle topluma potansiyel tehditleri tahmin edebilir. Bu çalışmanın amacı, büyük 5 kişiliği tahmin etmek için Facebook Kullanıcı veri ve aktivite kullanan bir öngörü modeli inşa etmektir. Bunu yapmak için, bu çalışma, beğeni, etkinlik, grup, etiket, güncelleme, ağ boyutu, ilişki durumu, yaş ve cinsiyet sayısı olan üç farklı çalışmada vurgulanan özellikleri bir araya getirmektedir. Çalışma myPersonality veritabanından alınan 7438 benzersiz Facebook katılımcısı üzerinde yapıldı. Bu çalışmanın bulguları, bir kişinin kişiliği sadece Facebook aktivitesini analiz ederek tahmin edilebilir ne kadar gösterdi. Ann modeli doğru bir 85% tahmin doğruluğu ile bireyin kişiliği sınıflandırmak başardı. Bu çalışmada üç farklı çalışmada türetilen özellikleri birleştirerek bir model öneriyor ve kelime veya durum güncellemeleri içeriği dahil olmadan tek başına bu özelliklere dayalı kişilik tahmin.

Anahtar kelimeler: Büyük 5 Kişilik; Facebook; Kişilik tahmini; Makine öğrenme; Yapay sinir ağı;

TABLE OF CONTENTS

ACKNOWLEDGMENTS	i
ABSTRACT	ii
ÖZET	iii
LIST OF TABLES	i
LIST OF FIGURES	ii
LIST OF ABBREVIATIONS	iii
CHAPTER 1: INTRODUCTION.....	1
1.1 Background.....	1
1.2 The Problem.....	3
1.3 Aim of the Study	4
1.4 Significance of the Study	4
1.5 The Limitations of the Study	5
1.6 Overview of the Study	5
CHAPTER 2: LITERATURE REVIEW	7
2.1 Big 5 Personality.....	7
2.2 Multi-label Classification.....	8
2.3 Artificial Neural Network	10
2.4 Related Studies.....	11
2.4.1 Using ANN for Prediction	11
2.4.2 Using ANN for Multi-Label Classification.....	13
2.4.3 Personality Prediction through Social Media.....	14
CHAPTER 3: THEORETICAL FRAMEWORK	17
3.1 Artificial Neural Networks.....	17
3.2 Multi-layer perceptron model	17
3.2.1 Back Propagation Supervised Learning	20
3.2.2 Optimization	20
3.2.3 Regularization and overfitting	22
3.3 ANN on Multi-Label Classification.....	23
3.4 Achievements of ANN.....	23
3.5 Strength of ANN	24

CHAPTER 4: METHODOLOGY	26
4.1 Model Development.....	26
4.2 Algorithm.....	28
4.3 Data and Pre processing.....	28
4.4 Transformation.....	31
4.5 Classification Architecture.....	32
4.6 ANN Multi-Label Classification.....	33
4.7 Keras and Tensorflow	35
4.8 Training, Testing and Validation	37
4.9 Visualization	38
CHAPTER 5: RESULTS AND DISCUSSION	39
5.1 Experimental Setup	39
5.2 Training and Testing	42
5.2.1 Training.....	42
5.2.2 Testing.....	45
CHAPTER 6: CONCLUSION AND RECOMMENDATIONS	48
6.1 Conclusion	48
REFERENCES.....	50
APPENDIX.....	59
SOURCE CODE.....	59

LIST OF TABLES

Table 2.1: Big 5 Personality traits dimension

Table 2.2: Example of MLC Problem

Table 4.1: Big 5 Personality Distribution

Table 5.1: Back propagation neural network training parameter

Table 5.2: Back propagation neural network training and testing results

LIST OF FIGURES

Figure 2.1: Simple Neural Network

Figure 3.1: Feed Forward Neural Network

Figure 3.2: “s-shaped” curved produced by the sigmoid function restricted to 0 and 1

Figure 3.3: Different Optimization Functions

Figure 3.4: Overfitting and Underfitting

Figure 4.1: Predictive Model

Figure 4.2: Model Process

Figure 4.3: Distribution by Gender

Figure 4.4: Distribution by Age

Figure 4.5: Neural Network model

Figure 4.6: The Flow Diagram for study framework

Figure 4.7: A Tensorflow Dataflow Diagram

Figure 4.8: 4 Fold Cross Validation

Figure 5.1: Before OneHotEncoding

Figure 5.2: After OneHotEncoding

Figure 5.3: Sample of input data after transformation

Figure 5.4: Sample of Transformed Output into binary

Figure 5.5: Accuracy and Loss for Scheme 1 Test 1(75:25 Split)

Figure 5.6: Accuracy and Loss for Scheme 1 Test 2(67:33 split)

Figure 5.7: Accuracy and Loss for Scheme 2 Test 1(K-10 Fold)

Figure 5.8: Accuracy and Loss for Scheme 2 Test 2(K-5 Fold)

LIST OF ABBREVIATIONS

ANN:	Artificial Neural Network
BP:	Back Propagation
BPNN	Back Propagation Neural Network
BP-MLL	Back Propagation Multi-Label Learning
CNN	Convolutional Neural Networks
KNN	K Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator Algorithm
LIWC	Linguistic Inquiry and Word Count
LR	Linear Regression
ML:	Machine Learning
ML-KNN:	Multi label K Nearest Neighbors.
MLC	Multi-Label Classification
MLP	Multi-Layer Perceptron
NB	Naïve Bayes
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error
SPSS	Statistical Package for the Social Sciences

CHAPTER 1

INTRODUCTION

1.1 Background

Over the last two decades social media and its prevalent use has become an integral part of our lives. The way people express opinions and sentiments has greatly changed due to social networking. Each of these social media sites; Academia, Facebook, Instagram, etc. are based on the concept of getting its users to share their experiences, opinions and various moments of their lives. A voluminous amount of data is constantly being exchanged on these social media sites everyday containing massive amount of interactive data.

A lot of people share a lot about themselves, their photos, videos and activities on these platforms so social media sites actually affects our real life. For example, twitter and Facebook has become a great avenue to share news and information.

Due to the massive information on social networking site, it has caught the attention of many researchers. Researchers have come to understand that with the volume of information obtainable from this social networking site, it can reveal a lot about human behavior and social interactions. Facebook is the social networking site with the highest amount of attention from researchers because it has the highest amount of active subscribers having over 2 billion subscribers (Statista, 2018) and has a lot of personal information (Wilson et al 2012).

With so much information constantly, being exchanged everyday on Facebook; this has made it possible for the prediction of various attributes just by looking at Facebook footprints. Some of these features include predicting the future (Asur and Huberman, 2010), predicting friendship ties with social media (Gilbert and Karahalios, 2009), predicting the stock market (Nguyen, Shirai, and Velcin, 2015) and many more.

Among these, predicting personality from various social media traits has become popular. With so many users active on Facebook and with the amount of information exchanged everyday by these users, it allows researchers analyze these data to understand the different personality traits of these users. The actual personality of a user can be gotten from the Facebook profile of a

particular user, thereby implying that by analyzing a person's Facebook activity and information, the personality of that individual can be extracted (Back et al., 2010).

Different techniques have been applied so far in literature and various studies have shown that there is a clear link between an individual and their Facebook profile, this link can be harnessed and applied in different areas such as targeted marketing, psychology and more. (Golbeck, Robles and Turner, 2011)

Using Facebook data to determine a person's personality trait based upon the big 5 personality model can be classified as a "multi-label classification" (MLC) problem, in the sense that an individual can possess more than one personality trait. Each of these five personality traits all corresponds to a classifier. An MLC problem is a problem where more than one target label is attached to each instance. This method is mostly applied to task such as text categorization, medical diagnosis, music categorization and semantic scene classification (Tsoumakas and Ioannis, 2006). In the big 5 model of personality, individuals differ in terms of openness, conscientiousness, extraversion, agreeableness and neuroticism (OCEAN) (Costa and McCrae, 1992), an individual can be categorized under more than one personality, for this reason the problem is called a MLC problem.

Predicting outcomes in an MLC problem can be seen as a complex problem and requires a model that is better in handling more complex and practical problems.

Different techniques have been proposed to solve problems such as these, some of which are; ML-KNN (M.L. Zhang and Zhou, 2007), Artificial Neural Network (ANN), Naïve Bayes, support vector machine (SVM) Decision Trees and Logistic Regression (Hall, 2017). ANN is a type of multi-dimensional regression analysis model, which makes it in various ways better than other regression models. The inspiration behind the development of ANN is stemmed on developing an intelligent system that can perform task intelligently like the human brain (Devi, Reddy, and Kumar, 2012). Regardless of how complex a system might be, ANN can accurately perform prediction problems, this is why a lot of researchers use it for prediction problem especially in cases where the problem is a too complicated to express in a mathematical formula and also in a case where the input/output data is available (Bataineh, Abdel-Malek, and Marler, 2012).

This study aims to use ANN to predict personality with data derived from Facebook data. Some studies use linguistic behavior of a person from a person's status update to predict personality (Tandera et al., 2017) but this study seeks to predict personality by analyzing and utilizing the relationship between a user's personality and their Facebook activities. The back propagation algorithm for neural network was used but since the data to be analyzed is a multi-label classification problem, some important characteristics of multi-label learning are not captured with the basic BP algorithm, which does not consider correlations of different labels. A modified BP algorithm better suited for ML problems was used. There are significant relationships between an individual's personality and their Facebook activity, this is to say that based on a person's Facebook activity one can get clues to a person's personality (Sumner, Byers, and Shearing, 2011). This study investigates to see if the similarities between an individual's personality and their Facebook activity can be used to better predict personality more successfully.

1.2 The Problem

Nowadays Social media has become an integral part of our daily lives. A lot of personal information is constantly being uploaded on Facebook. In a recent article by Auchard and Ingram (2018) speaks on how Facebook data was used to target voters during the 2016 United States election and manipulate the election. This goes to show that so much can be discovered about individuals on Facebook just by analyzing Facebook data. With this in mind it's obvious to ask what more can be derived from this data, that is why personality prediction has become an important aspect of social media. There is a significant correlation between personality and Facebook activity such as number of likes, tags, status updates, friends, events. Although many researches have been carried in the area of social media and personality, not so much has been done in harnessing this information for businesses, crime and more.

Being able to use Facebook data to understand the personality of the users, businesses can harness the information to better expand their business and reach their target market. People with a high tendency to commit crimes can be easily predicted using Facebook data and people can also know the personality of people before going into any relationship with them. Neural network is rapidly growing as an interesting tool for building predictive models especial for solving complex problems. This study intends to investigate linkage between a user's Facebook

activity and their personality by using a neural network predictive model to analyze information gotten from the users Facebook activity. This will help to know the extent of relationship and to know if this can help better predict a user's personality more accurately.

1.3 Aim of the Study

The aim of this study is to understand the extent as to which the personality of an individual can be inferred from their Facebook activities, and in order to accomplish this, it is important to address the following research questions below:

- Are user activities, network information influential factors in predicting the personality of Facebook users moderated by gender and age of Facebook users?
- How should these factors be presented in other to derive accurate predictive patterns for personality prediction?
- How can neural networks be trained so as to learn predictive patterns for personality predictions?

1.4 Significance of the Study

Facebook consist of over 2 billion active users making about one third of the world's population, developing a model with a high accuracy in personality prediction can go a long way in the business sector, education, relationship, law enforcement and much more, thereby making this study a relevant information system research. Machine learning in computer information systems helps business and public organizations provide the necessary expert and intelligent systems required to help with decision making process in a constantly evolving field. Currently some companies such as Timber and eHarmony are constantly working to improve online dating with machine learning and some features which include the big 5 personalities (Chowdhury, 2017), a predictive model that can accurately predict personality just from Facebook activity can go a long way in online dating. In Adaptive systems, user modelling is very essential. Understanding the goal of an adaptive system in respect to some of the user features can go a long way in proper serving the user (Kobsa, 2007) and one interesting user feature to consider is personality. Understanding a user's personality can help identify some variables such as needs in different context. A model that can accurately predict personality may help adaptive applications adapt to user's behavior accordingly. For example, in e-commerce products can be offered to users can

vary depending on their personality with respect to Impulsive sensation seeking (Ortigosa, Carro, & Quiroga, 2014). The personality of an individual is stable through time and situation (Espinosa and Rodríguez, 2004), meaning personality of an individual doesn't change online or offline, an individual that is sociable offline will be sociable online. Therefore, the Facebook profile of an individual can reflect actual personality (Back et al., 2010). There are some studies in literature that predicts big 5 personality utilizing features such as linguistic which is retrieved from written text or speech text (Mohammad and Kiritchenko, 2013), However the topic of predicting personality on social media has become a popular one. The pacesetter well known research was by (Golbeck et al., 2011). There are some other studies that employs linguistic inquiry and word count (LIWC) (Sumner et al., 2011) , structured programming for linguistic cue extraction(SPLICE) (Tandera et al., 2017), time related features (Farnadi, Zoghbi, Moens, & Cock, 2013) and others. This study contributes to an expanding literature on inferring personality with social media by using back forward feed forward algorithm to analyze the Facebook activity data in order to see if better prediction results can be achieved. As at the time of this study, there is no knowledge of any literature that uses neural network strictly together with Facebook activity without looking at post and text to predict personality. Also, other current studies available uses a small data set for analysis which might impede the reliability of the results, this thesis analyzed dataset retrieved from myPersonality database (Kosinski et al., 2015) which consist of over 3 million Facebook users.

1.5 The Limitations of the Study

In regardless of the fact that this study will attain its goal, some restrictions that are attached to it still exist due to some factors.

- Some amount of data was excluded from analysis due to missing data in some columns
- Study dataset is limited to Facebook data

1.6 Overview of the Study

The study is made up of six chapters in all:

Chapter 1 gives a general insight on social media, the big five model, neural networks, the issues, definition, the extent of the study, the importance of the study, the limitations of this study and finally the breakdown of this study.

Chapter 2 Introduces the related topics and studies to this study and gives a brief introduction to Artificial Neural Network and multi-label classification

Chapter 3 outlines the hypothetical systems, how ANN works, the different underlying factors that makes up ANN and its foundation, its benefits and so on.

Chapter 4 Presents the details of the instrumentation, tools and models used for this study and the philosophy behind their implementations.

Chapter 5 discusses the outcomes and experiments conducted in this study

Chapter 6 Finalizes the study, restates importance and gives future recommendations for study.

CHAPTER 2

LITERATURE REVIEW

In this chapter, a brief explanation about the big 5 personality and its facets was presented, A brief back ground on multi label classification, a brief background on neural network and finally different studies previously published in this subject area were examined and analyzed.

2.1 Big 5 Personality

In In psychology, there are five major characteristics that define human personality known as “big 5”, this is a well experimented and scrutinized structured for individual personality used by researchers recently (Goldberg, 1992). This big 5 personality trait is divided into Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism. Over the years, this big 5 models have become standard for personality due to the fact that it came out of prior test on personality, and the test also showed that the models validity was not altered by languages or variation in method analysis (McCrae & John, 1992), therefore resulting in its acceptance. Below is a detailed explanation of the big 5 personality;

- Openness: Intelligent, curious and open to new things and ideas: Appreciate diverse views, experiences and very imaginative (Lima & de Castro, 2014)
- Conscientiousness: Extremely reliable, task oriented and well-organized people. They ensure to complete every task. They tend to commit themselves to their work, they plan ahead and very responsible (Adali, Sisenda, and Magdon-Ismail, 2012)
- Extraversion: Energetic, Friendly, enthusiastic and attractive to people. They are outgoing and quick to make friends. They also exhibit traits of peace making it easy to get along with people (S. Adali and Golbeck, 2012)
- Agreeableness: Exhibits optimism traits, calm, peace keepers, trusting and nurturing with a high tendency of trying to help others (S. Adali and Golbeck, 2012)

- Neuroticism: High traits of insecurity, not so good with others, very sensitive; that is to say, they easily get affected with negative emotions. (S. Adali and Golbeck, 2012).

Table 2.1: Big 5 Personality traits dimension (Ateş, 2014)

Openness	Conscientious	Extroversion	Agreeableness	Neuroticism
Imaginative,	Organized,	Energetic,	Sympathetic,	Anxious,
Wide interest,	Disciplined,	Forceful,	Straight	Tense,
Curious,	Planner, Goal	Adventurous,	forward,	Worried,
Intelligent,	oriented, not	Enthusiastic	Compliance,	irritable,
Artistic,	impulsive		Generous	impulsive, shy
Unconventional				

When dealing with the big 5 personality model, each individual can highly exhibit some of these traits together therefore meaning that the personality traits are not opposed to each other. A person can exhibit high symptoms of Agreeableness, Openness, while exhibiting little symptoms of Neuroticism.

2.2 Multi-label Classification

The big 5 personality traits are independent of one another; an individual can exhibit high symptoms of more than one personality trait hence making it a multi-label learning task.

In machine learning, multi-label classification (MLC) is a form of classification problems but varies differently from other classification problems, in the sense that each sample can have several labels (Tsoumakas and Ioannis, 2006). This varies from other classification problem that can have just one label and never two (i.e. an object can either be classified as dog or cat but never both) this is known as Multi-Class Classification. In MLC samples are attempted to be classified in more than one label (that is a person be both labelled as openness and agreeableness) (Tsoumakas and Ioannis, 2006). There are various real-world situations where MLC can be applied such as classifying a movie genre which can be both comedy and action.

The method of solving MLC problems can be grouped into two; problem transformation and algorithm adaptation.

Algorithm Adaptation uses algorithm to directly alter and classify standard classification technique to perform MLC. This schema treats MLC as a single integrated problem without requiring problem transformation. Some examples of machine learning methods that have adapted this approach in handling MLC are; ANN, boosting, decision trees and KNN (Hall, 2017).

The problem transformation method transforms the problem into a series of simpler bitwise classification problems and two tactics are used for transformation, binary relevance and label powerset (Read et al., 2011).

Binary relevance is the baseline method when using problem transformation method, for each label it independently trains one binary classifier. One can look at this transformation method as an extension of a binary classifier applied in a one-vs-all method, that is, each task is labelled as either 1 or 0, present or absent (Read et al., 2011).

In the label powerset transformation method, the numbers of labels are expanded by creating one binary classifier per label combination which is certified in the training data set (Tsoumakas & Ioannis, 2006). For both binary relevance and label powerset some algorithms such as SVM, Naïve Bayes, K Nearest Neighbours has been used in this method (Read et al., 2011).

Table 2.2: Example of MLC Problem

Input Variables					Output Variables			
X ₁	X ₂	X ₃	X ₄	X ₅	Y ₁	Y ₂	Y ₃	Y ₄
1	0.3	0.5	1	0	1	1	1	0
1	0.7	0.2	1	1	1	1	0	0
0	0.2	0.3	0	1	0	1	1	0
1	0.4	0.7	0	0	1	1	0	1
0	0.6	0.6	0	1	0	1	1	1
0	0.4	0.4	1	1	?	?	?	?

2.3 Artificial Neural Network

ANN are designed to work as the biological nervous systems works in interacting with objects of the real world, they are a large parallel interconnected networks made up of nodes and each node is referred to as neurons (Zhang and Zhou, 2006). ANN has the ability to learn, to adapt by modifying its internal structure depending on the data that passes through it. It is one of the most successful learning methods and has performed so well in classification (J. Zhang, 2016). ANN provides variations of techniques to learn from examples and performs very well in pattern recognition. At the moment various types of neural networks exist such as self-organizing feature mapping networks, radial basis function networks, adaptive resonance theory models and of course multi-layer feed forward neural networks (Kalghatgi et al., 2015).

ANN can be distinguished based on the strategy used for learning, there are two major learning strategies used for learning in ANN; supervised and unsupervised

- **Supervised learning:** In supervised learning, the network is given both the input and output data, understanding that there is a relationship between the input and output, it adjusts its weight to try to produce the same result with the output based on the different scenario it has been fed with (Lison, 2012).
- **Unsupervised learning:** In unsupervised learning the output is not known by the network, only the input is given. The network tries to recognize patterns based on these inputs it received and groups same patterns as clusters (Lison, 2012).

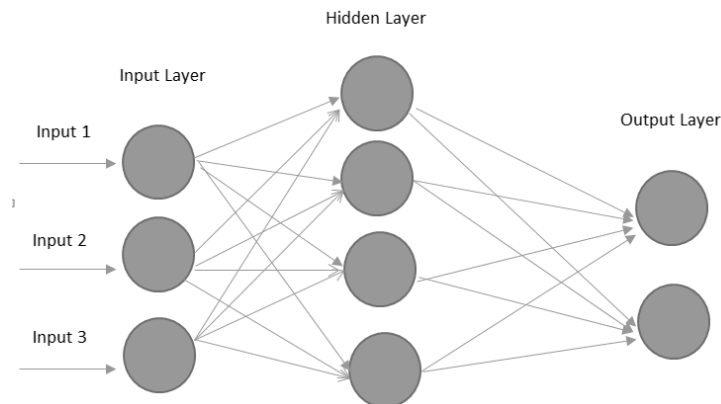


Figure 2.1: Simple Neural Network (Kalghatgi et al., 2015)

2.4 Related Studies

A lot of studies has been carried out in the past using ANN as a tool, in this section studies carried out using ANN for prediction were examined, then after that studies carried out in the area of ANN in prediction for multi-label classification problems were examined and then finally studies relating to ANN in personality prediction.

2.4.1 Using ANN for Prediction

Different models and methods have been proposed for prediction of various outcomes. In 2010 ANN was used as a tool to predict team performance by analyzing individual past achievements and history (Hedberg et al., 2010). The aim of the study was to provide a means by which employers can analyze prospective team member's track record to understand the effect of that individual in the team. After analysis, training, testing and evaluation, the model achieved 73.4% prediction accuracy. With this level of accuracy, the study claims that this ANN approach can be applied in other organizational levels including recruitment.

Champa and AnandaKumar, (2010) study was on human behavior prediction through handwriting analysis. The study uses ANN to analyze various samples of individual handwriting by looking at the baseline, the pen pressure and the letter 't'. The study states that professional handwriting examiners can understand human personality from and individual's handwriting however the process is costly and prone to fatigue. The baseline, the pen pressure and the height of the of the t-bar in the letter 't' stem were fed into the ANN as input and outputs individual personality trait. The model was run through various epochs and hidden layer and attained a maximum accuracy of 53%.

Another study by Nkoana, 2011 proposes an ANN model for flood prediction and early warning, in the study various number of trained neural network architectures were evaluated using their mean percentage accuracy. The study implemented 14 neural networks using daily rainfall as the predictive variable from the period of 1995 to 2009, after examining the performance of the neural networks the Elman recurrent neural network with two hidden layers and two hidden nodes yielded a better result of 58% accuracy. The study claims that using ANN with daily

rainfall can be used to predict floods. Another study by Devi et al., 2012 also proposes an ANN model for Weather prediction. The study collects data from atmospheric pressure, temperature, wind speed, wind direction, humidity and precipitation and uses it to train a three-layer ANN. The results were compared with practical working of the meteorological department and the study claims to have built a model which can successful predict weather based on the comparison results.

Another interesting study using ANN for future predictions was by Song and Kim (2014), the study feeds the big five personality trait as input into the ANN model to predict individuals future location. The study exploits the connection between human mobility patterns and their personality to train the ANN to predict future locations. The study combined time information and personality as input nodes while locations as output sample training data. The study claims to have been able to predict human location through the help of the personality trait. The study recommends and inverse of this model in the future to use mobility pattern to predict personality.

Binh and Duy (2017) uses ANN as a tool to predict student performance based on the students learning style. The study conducted an online survey with a participation of 316 undergraduate students in various courses. Using the data collected and analyzed an ANN model was built to predict students' performance based on their learning style. The ANN model managed to produce 80.63% classification accuracy, the study claims that this can method can be applied in e-learning environment adaptive models that can support learners.

Al-Shihi et al., (2018) proposes a model that can be used to predict mobile learning adoption in developing countries. The study integrates some constructs such as social learning, flexibility learning, enjoyment learning and economic learning. The study was conducted on 388 participants from major universities/colleges at Oman and ANN was used as the tool for prediction. The study claims that this model can be used to predict and influence mobile learning adoption.

2.4.2 Using ANN for Multi-Label Classification

Nam et al. (2014) proposed a simpler ANN approach to handle multi label classification in largescale multi-label text classification. The proposed method is aimed at being an alternative and better method than the state of the art back propagation multi label learning approach. In the study the BP-MLL's pairwise ranking loss was replaced with cross entropy also, and other features such as ReLU activation function was used together with AdaGrad optimizers.

The study claims that this approach enables the model converge in just a few steps and the dropouts utilized helps prevent overfitting. The study evaluates the performance of the proposed model with other baseline models. The algorithm trains with a higher convergence speed due to the ReLU activation, the model also uses dropout to prevent overfitting by randomly dropping individual hidden units while by taking advantage of label space inherent correlation to minimize rank loss.

In 2015 Liu and Chen proposed a multi-label approach for sentiment analysis of microblogs. The study compares 11 state of the art ML classification methods and uses 8 metrics for evaluation. The comparison was carried out on 2 microblog datasets. Out of the 11 methods evaluated, some of the methods performed better than others depending on the scenario. Rakel (Random K label set) performs better with HR, while other algorithms performed better on AI. So, the different features in the results affected the results of the study but the result of the study shows that one of the dictionaries used in the study Dalian University of Technology Sentiment Dictionary with homer performs best on multi-label classification.

In 2016 Corani and Scanagatta proposed a multi-label classifier model which is based on Bayesian networks but performs slightly different from the baseline Bayesian network. The model addresses the dependencies amongst the class variable which is normally overlooked when devising independent classifier for each of the classes to be predicted. The model works by simultaneously predicting the class variable which is different from the baseline approach, the study result show that the performance of the proposed model out performs the independent approach when predicting multiple air pollutions.

Another study by Tabatabaeiet al. (2017) examines two different (Random K-label sets and multi-label K-Nearest neighbours) multi-label classification method and proposes a model to disaggregate appliances in a power signal, after which the study evaluates the model on different real world scenario. The study claims that the evaluation results carried out by comparing with existing literature shows that the proposed classifier were competitive with existing literature.

Still in 2017 Kee et al. proposes a neural network multi-label classification system to predict the arrival time of bus transport. The neural network is built based on the historical GPS (Global Positions System) arrival time and ensemble of neural network is used to improve the reliability of the output. The results of the study show that the proposed model is able to forecast the arrival time up to a reasonable percentage of 75%. The neural network and ensemble model was compared with other algorithms such as decision tree, Random forest, Naïve Bayes, and the model proves to be 8% better than the other algorithm. The study suggests further improvement of the model by using power transformation and some other different ensemble methods

2.4.3 Personality Prediction through Social Media

In 2012, Wald, et al. proposes a form of machine learning ensemble learning called SelectRUSBoost to predict psychopathy through twitter data; this method adds feature selection an imbalance aware ensemble to tackle high dimensionality. The study states that when ensemble learning, data sampling and feature selection in SelectRUSBoost, the model is able to hit AUC (Area under the curve) of 0.736 and this performance is only achieved when this model is used. The study states that a model such as this can be used by law enforcement in discovering psychopathic states through their twitter data. The study also states that though the model can be used with twitter to predict the incidence of psychopath they are not sufficient to provoke direct actions but can be used to flag potential risk.

Farnadi et al. (2013) explores the use of machine learning (SVM, NB, KNN) to infer personality just by examining Facebook status updates of various users. The study strengthens their prediction model by not just relying on one source but by including different training samples from another source (Essay corpus) helping the study show that trait can be generalized across social media platforms. The study investigates 250 users with 9917 status updates and states that despite having a small amount of dataset the model could still outperform other baseline

methods. Another study by Kandias et al. (2013) proposes a methodology that detects users that are hostile or with a negative attitude towards the authorities, the study combines the dictionary learning based approach and machine learning techniques (SVM, NB, LR). The study analyzed information posted on the YouTube website

Lima and de Castro (2014) study uses a semi supervised classification approach to predict personality through twitter data. The data takes a different approach from other studies, this study doesn't take user profile into consideration and it doesn't work with single texts like in other studies but works with a group of text. The study uses the problem transformation method to transform the problem into five binary classification problems. The study used three well established machine learning algorithm; NB, MLP and SVM to train the proposed system and was applied to predict personality from tweets which resulted in an 83% prediction accuracy.

Kalghatgi et al., (2015) also investigates big 5 personality trait prediction through analyzing tweeter data with ANN. The study explores the parallelism between an individual's linguistic information and their big five personality trait and uses the tweets posted by an individual to predict personality. The study also says that the model doesn't take user tweeter profile into consideration and implements it in java NetBeans using Hadoop framework to make predictions of multiple individuals at the same time.

In 2016, Akshat investigates using CNN to predict personality from social media images, the study sort to find out if there was any relationship between the output why such relationship exists. The study results show how powerful Neural network is as a tool to measuring and learning highly non-linear mappings between input data and output data. The study uses the transformation method to transform the task into a classification task and uses a chance baseline which guesses just the highest occurring class which is used for comparison. The model was trained and validated with a split of 80, 10, and 10 for training, testing and validation. Another study by Li et al. (2016) extracts emoticon features and linguistic features from Facebook data and uses it to predict the big five personality trait, the also strengthens the robustness of their model by applying cross-domain learning algorithm and features. The study implements ANN, LR and M5P as algorithms and Root Mean Square Error (RMSE) as the standard for evaluation. The study claims that the model shows better performance than results in other literature.

In 2017, Tandra et al. (2017) carries out a competitive analysis of current deep learning architecture and uses accuracy results to compare performance. The study involved using the models to predict big five personality trait from data retrieved from users Facebook account. The dataset used in the study were gotten from two different sources; myPersonality dataset consisting of 250 users and then 150 Facebook user's data which were collected manually. The study also uses linguistic features such as LIWC with both closed and open vocabulary approach. The study reports saying the model outperforms other methods by 74.14% average accuracy, though accuracy was low with some traits, study claims this could be a result of limited dataset. The experiment results show ANN doing better than other traditional machine learning classification method. Again in 2017, Laleh and Shahram proposes a model that uses LASSO algorithm to select the best features and predict the big five personality trait from a user's Facebook data by examining Facebook likes. The study examines the likes of 92225 users while combining with 600 weighted topics, the model also examines the task as a regression problem. The training and test data is split 75% and 25%. The cross-validation method was used to validate the model. Still in 2017 is a study by Iatan which uses Fuzzy Gaussian Neural Network (FGNN) to predict personality from a user's Facebook account based on the data publicly available and compares result with two other models; multiple linear regression model and multi-layer perceptron. The performance of the model was tested using normalized root mean square. The study results show how the proposed method outperforms the other two methods both during and training

CHAPTER 3

THEORETICAL FRAMEWORK

3.1 Artificial Neural Networks

Artificial Neural network is a sub field of machine learning; this involves the learning of representations derived from data with emphasis on learning successive layering of related meaningful representation. In most cases deep learning is often referred to as ANN but in as much as ANN and deep learning could be seen as one as the same, deep learning are often identified by their increase in layers and complexity in structure, where as a basic ANN can have just one single hidden layer. The name ‘deep’ is derived from the increase and piles up of layers in the model, the more the layer the deeper the model (Chollet, 2017). Since learning involves learning nonlinearity from samples, ANN helps improve representation capability. The nonlinearity’s form can be learnt from just a simple algorithm (J. Zhang, 2016). When dealing with this layered representation, the model that are almost always used is for learning are artificial neural networks (Chollet, 2017).

ANN draws its central concepts from the brain in that just as successive neurons respond to stimuli so also ANN are organized into layers which responds to input by further stimulating the next layer. All other machine learning can be described as learning through past observations to make predictions, ANN on the other hand doesn’t just make predictions but learns to correctly represent and map the data. So, in summary ANN is about mapping inputs to target outputs which is actually done by the model observing several mappings of inputs to target through a sequence of layering or data transformation.

3.2 Multi-layer perceptron model

The multi-layer perceptron is a feed forward ANN model that takes several inputs which has an associating weighing factor and produces an output. With these weights which essentially are a bunch of numbers intertwines with the input respectively to contribute to a different degree in which the output is expressed. This output is determined by checking if the weighted sum is greater than some certain threshold set by the network bias, if the weighted sum is greater it

assigns it as 1 but it not it assigns it as 0. Basically, learning in ANN is finding a set of values for the weights so that the network can accurately map samples inputs to their respective outputs (Schmidhuber, 2015).

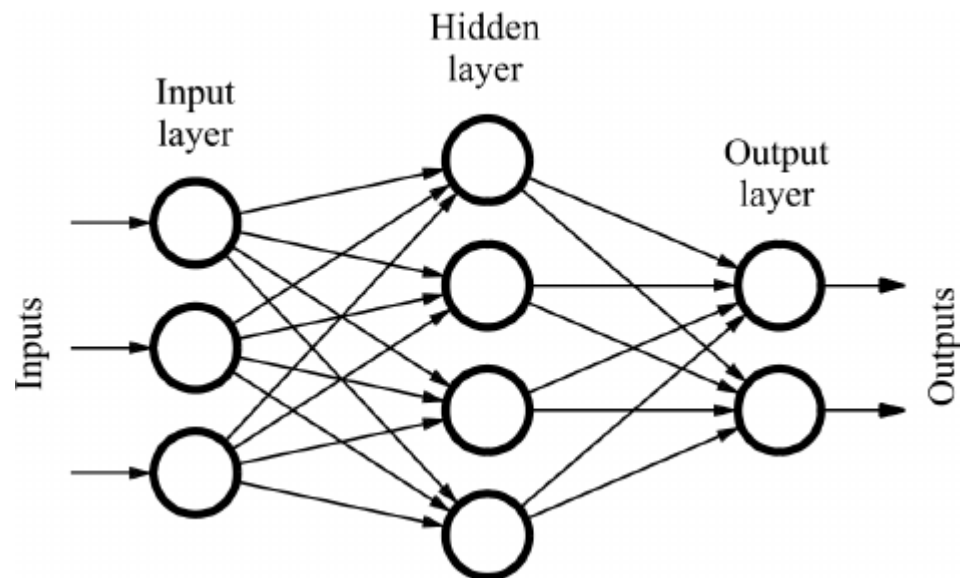


Figure 3.1: Multi-Layer Perceptron Feed Forward Neural Network Model

In the diagram in figure 3.1, the first layer is the input layer with n number of inputs, the second layer is the hidden layer with 4 neurons and the third layer is the output layer with m output neurons.

One thing about ANN is that some models can have thousands and millions of parameters, finding the right value to fit all might be a very daunting operation. The network value can be easily altered by a little shift in the weight or a little change in the bias that can easily make the network flip completely and altering results tremendously (Nielsen, 2015). Manually adjusting these weights without flipping up the network completely can be tricky but this is overcome by the neuron called the sigmoid neuron. The sigmoid neuron consists of a weighted input which takes a value between 0 to 1 this then produces an output which is similar to the networks perceptron. This transformation into binary is defined by the sigmoid function.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.1)$$

The benefit of the sigmoid function is that it makes the ANN more robust and resilient against minor changes which allows for fine tuning of the network without completely flipping the behavior of the network (Harrington, 2012). The equation below represents figure 3.1 and illustrates a feed forward ANN with its weights (w) and activation functions (f)

$$y_m = \hat{f} \left(\sum_{j=0}^m w_{j4}^{(2)} f \left(\sum_{i=0}^n w_{4i}^{(1)} x_i \right) \right) \quad (3.2)$$

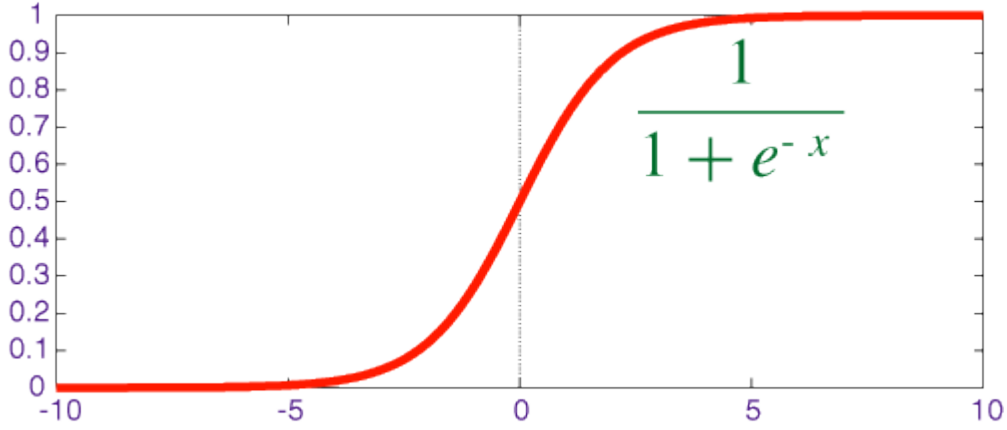


Figure 3.2: This is a characteristic “s-shaped” curved produced by the sigmoid function restricted to 0 and1 (Harrington, 2012)

That being said it is important to keep track of how well the network is performing, it is important to measure how far the output produced is from the expected output and this is where loss function comes in. the job of the loss function is to compute a distance score between the predictions and the target to determine how well or how off the network had performed. The loss is a summation of the errors made for each example in training or validations sets. The trick here is using the score as a feedback to adjust the weights in a direction that will reduce the loss score. In classification the is usually negative log-likelihood. The optimizer is what carries out the job of adjusting these weights through a supervised learning technique which is the central algorithm in ANN known as back propagation.

3.2.1 Back Propagation Supervised Learning

Backpropagation (short for backward propagation of errors) sometimes also referred to as reverse-mode differentiation, is the most widely utilized algorithm for adjusting ANN parameters. In this method, the configuration is set and the data is presented to the ANN. Back propagation is in two faces that are forward pass and backward pass. In the forward pass, the data is fed into the network, the result of presenting these data most likely outputs incorrect results that is handled by the loss function, the results are then retrieved from the output layer and propagated back to the input known as the backward pass to do the process all over again. As this happens, the weights are adjusted to reduce the error between the target output and the resulting output from training (McGonagle et al., 2017). At the end of a successful backpropagation learning there should be a smooth mapping of inputs to outputs which should be demonstrated by the internal parameters. Equation 3.3 shows a cost function for back propagation neural network, binary cross-entropy (BCE) loss function.

$$BCE = -\frac{1}{N} \sum_{i=1}^N y_i \log(h_{\theta}(x_i)) + (1 - y_i) \log(1 - h_{\theta}(x_i)) \quad (3.3)$$

The binary cross-entropy loss function used with binary classification when assumed that the output layer is transformed using the sigmoid activation function.

Assuming a set of target outputs label 0 and 1, the network is trained to maximize the log conditional probability within every given sample (x and y).

3.2.2 Optimization

Of all the different models that requires optimization, optimization in ANN is the most complicated (Goodfellow et al., 2016). Normally it can take days to months to solve just a little neural network problem, so to tackle this challenge various optimization techniques have been developed. Basically, the job of an optimizer is to enable the network update itself by adjusting its weights that is the internal parameters so as to reduce the cost function without flipping the entire model. Optimization algorithm can be divided into two; Constant learning rate algorithm and Adaptive learning rate algorithm.

In constant learning rate algorithm, the most commonly used is the stochastic gradient descent (SGD) which is a form of gradient descent. The SGD only randomly examines a subset of samples for calculating the gradient. The SGD obtains unbiased estimates of the gradient calculating the average gradient of the mini batches which has been randomly computed. The learning rate in SGD is its crucial parameter and choosing an adequate learning rate can be difficult. A learning rate that is too small leads to painful slow convergence, while a learning rate that is too large can hinder convergence (Goodfellow et al., 2016). The hyper parameter of gradient descent must be defined in advance and the type of model matters in the definition and this is a challenge. Adaptive gradient descent algorithms serve as an alternative to the classical SGD, some examples of adaptive gradient descent are Adam, Adagrad, Adadelata, RMSprop. These algorithms have per parameter learning rate methods that tunes hyper parameter without requiring expensive work or tuning manually. The difference between these algorithms lies in their computation power requirement and optimum result (Agrawal, 2017).

Usual SGD methods adapts updates to the slope of the models error function and then speed up the SGD but Adagrad on the other hand depending on importance, adapts updates to each individual parameter to perform larger updates for infrequent parameters or smaller updates for frequent parameters (Duchi, Hazan, and Singer, 2011). This makes it well suited for dealing with sparse data (Dean et al., 2012). The benefit of Adagrad is that it takes away the need to manually tune the learning rate. Adadelata is an extention of Adagrad which seeks to reduce monotonically decreasing learning rate, Adadelata restricts the window of accumulated past gradients to some fixed size. Another like Adadelata is RMSprop with the same purpose of dealing with Adagrad's diminishing learning rates. It divides the learning rate by an exponentially decaying average of squared gradients. Adam is a very popular method today; it computes the adaptive learning rate for each parameter and also keeps an exponentially decaying average of past gradients which is an addition to RMSprop and Adadelata that just stores an exponentially decaying average of past squared gradients (Kingma and Ba, 2015).

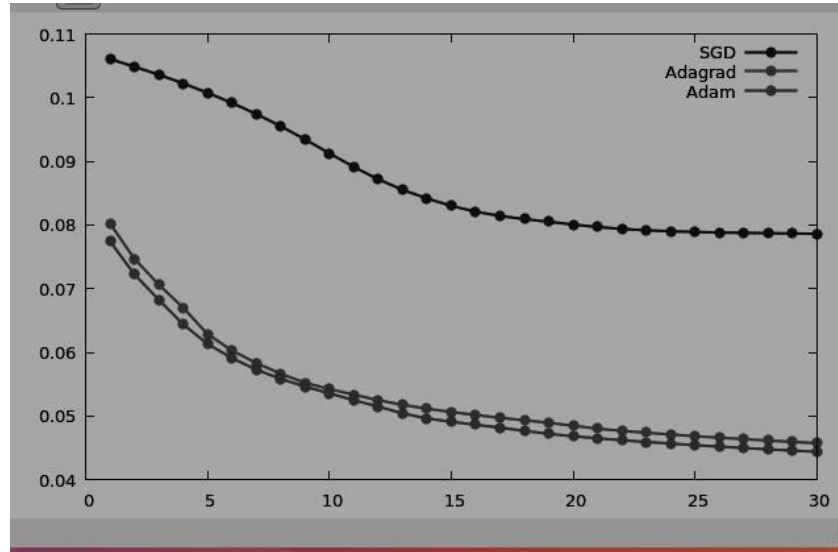


Figure 3.3: Different Optimization Functions (Agrawal, 2017)

3.2.3 Regularization and overfitting

When training ANN overfitting or under fitting is a normal phenomenon, when a model trains too well and gets too well fit to the training data, this is known as overfitting. The model has a high performance on the training data but performs very poorly with the test data. Overfitting happens in every ANN problem, learning how to deal with this is very crucial to mastering ANN (Chollet, 2017). On the other hand, when a model is not able to capture a sufficient low error value on the training data set due to the fact it does not fit the training data, this is known as under fitting (Goodfellow et al., 2016). Finding the balance between overfitting and under fitting that is; being able to come up with a model that isn't overfitting or under fitting is the challenge and one way to handle this issue is through regularization.

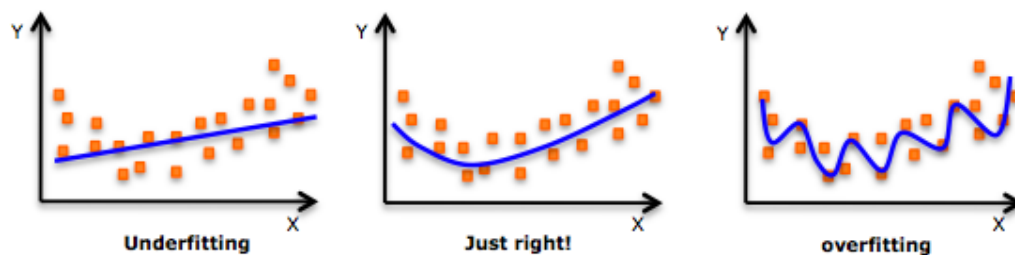


Figure 3.4: Overfitting and Underfitting (Bronshtein, 2017)

Regularization is any adjustment we make to a learning algorithm with the aim of not reducing its training error but its generalization error by adding a penalty term to it which helps to smoothen the decision boundary surface. Regularization is of a core importance in ANN, the only next thing that rivals it is optimization (Goodfellow et al., 2016).

3.3 ANN on Multi-Label Classification

Currently there are three variables driving the recent growth in ANN; Algorithmic advances, dataset and hardware. ANN has proven to be very well able to capture and model label dependencies in the output layer and it also performs excellently regardless of what data it receives as input. It has been able to show competitiveness not only in large data set but also in small dataset (W. Zhang et al., 2017). ANN also performs greatly with multi-label classification problems as a result of its algorithms which are constantly being improve and allows for better gradient propagation, some of this algorithm enabled; Enhanced activation functions, improved weight initialization and optimization (J. Liu et al., 2017). Also, in the ANN there are more advance techniques that which helps make learning in Multi-label classification problems more efficient, such as; residual connections, batch normalization depth wise separable convolutions (Chollet, 2017).

3.4 Achievements of ANN

ANN has been very crucial in the field of machine learning and has attained nothing short of revolution in this field. Things which come across to humans naturally but extremely impossible such as hearing and seeing, ANN have been able to achieve remarkable results in those areas. The following are some of the major achievements so far in ANN (Chollet, 2017);

- Recognizing speech almost like humans
- Transcribing handwriting almost like humans
- Classifying images almost like humans
- Great improvement in machine translation
- Improved conversion of text-to-speech
- Autonomous driving

- Better and improved search results
- Enhanced performance in answering natural language questions
- Content filtering on social media
- Recommendations on ecommerce websites

ANN continues to grow in its application on all sector business, health, government, art and is very much present in consumer products such as cameras and smartphones (LeCun et al., 2015). Even in cloud technology so many algorithm and methods involving ANN and constantly being designed to optimize and enhance the cloud service (Ejimogu and Başaran, 2017).

Although the achievements of ANN in the last few years have been remarkable, there are still so much grounds to cover in meeting the expectations placed but one thing is for sure ANN continually makes major breakthroughs in solving problems that has resisted the artificial intelligence community for some time now. As a matter of fact, deep learning in ANN has superseded the performance of other machine learning techniques in areas such as predictions in mutation for non-coding DNA on diseases and gene expression (Xiong et al., 2015), Analyzing data from particle accelerators (Ciodaro et al., 2012), predicting drug molecules potential effect (Ma et al., 2015), brain reconstruction (Helmstaedter et al., 2013) and so much more.

3.5 Strength of ANN

The reason why there is so much hype with ANN is because of the fact that it offered more satisfactory results on many problems (Chollet, 2017). Originally one of the most crucial step in machine learning workflow used to be feature engineering (Anderson et al., 2013) but ANN has made this a thing of the past because ANN completely automates the whole process (Kanter and Veeramachaneni, 2015). For example, some other machine learning approaches such as SVM and Decision tree only involved the transformation of the input data into probably one or two progressive representation spaces which can be called shallow learning, but the problem with this is that more complex problems cannot be properly handled with such techniques. This led humans going through great lengths to manually engineer the input data so as to be able to be processable by these techniques, this is what is known as feature engineering (Anderson et al., 2013). However, on the other ANN completely automates the process, humans don't need to go

through the stress of doing all these which hereby highly simplified the machine learning workflow.

In ANN data is not an issue, ANN can work directly on data regardless of it being an image, video or audio. The traditional machine learning algorithm processes the data in a particular way, humans have to tell the system what to look for in order for it to make a decision but in ANN, the algorithm does this thing on its own without necessarily being programmed to do so.

So many ANN libraries exist within the framework and these libraries are constantly growing, some examples which are multi-layered convolutional networks with back propagation which is perfect for image processing, multi-language processing and many more. Some popular deep learning libraries include MXNet (Chen et al., 2015), Caffe (Jia et al., 2014), Theano and Tensorflow (Abadi et al., 2016) which was used in this study

CHAPTER 4

METHODOLOGY

In this chapter, we discuss on the research methodology used for this study in predicting personality based on users Facebook activities. This chapter explains the various steps that were taken.

4.1 Model Development

Various studies have shown that Facebook users express themselves on Facebook as they do in real life and sometimes even more. The actual personality of a user can be gotten from the Facebook profile of a particular user, thereby implying that by analyzing a person's Facebook activity and information, the personality of that individual can be extracted (Back et al., 2010). A person who is an extrovert in real life always tends to post a lot about their activities and share their experiences while a person who is neurotic often tends to be less active and have less tags due to their shy nature.

As a result of the huge relationship between an individual's Facebook profile and their personality there are some predictive models that takes into account the Facebook activities of users and their networks (Bachrach et al., 2012). In the study features such as number of likes, groups, tags, friendship networks were the features focused on. Another study (Kosinski et al., 2013) also proposes a predictive model that just focuses on the demographic information retrieved from the users Facebook profile, demographics such as age, gender and relationship status.

All these previous works show a well tight relationship between the users Facebook profile regarding their usage pattern and their demographics. Based on this, in this study a different dataset is used and a combination of two different predictive models from previous work is used to formulate the model for this study.

The model used in the study was built on the findings of three studies. The model is a combination of the features highlighted by Bachrach et al. (2012) and Sumner et al. (2011) with the features highlighted by Kosinski et al. (2013) in their studies which are number of likes,

number of status updates, number of events, number of groups, number of tags, network size, relationship status, age and gender. In the studies combined these features were the features with a high influence on personality prediction.

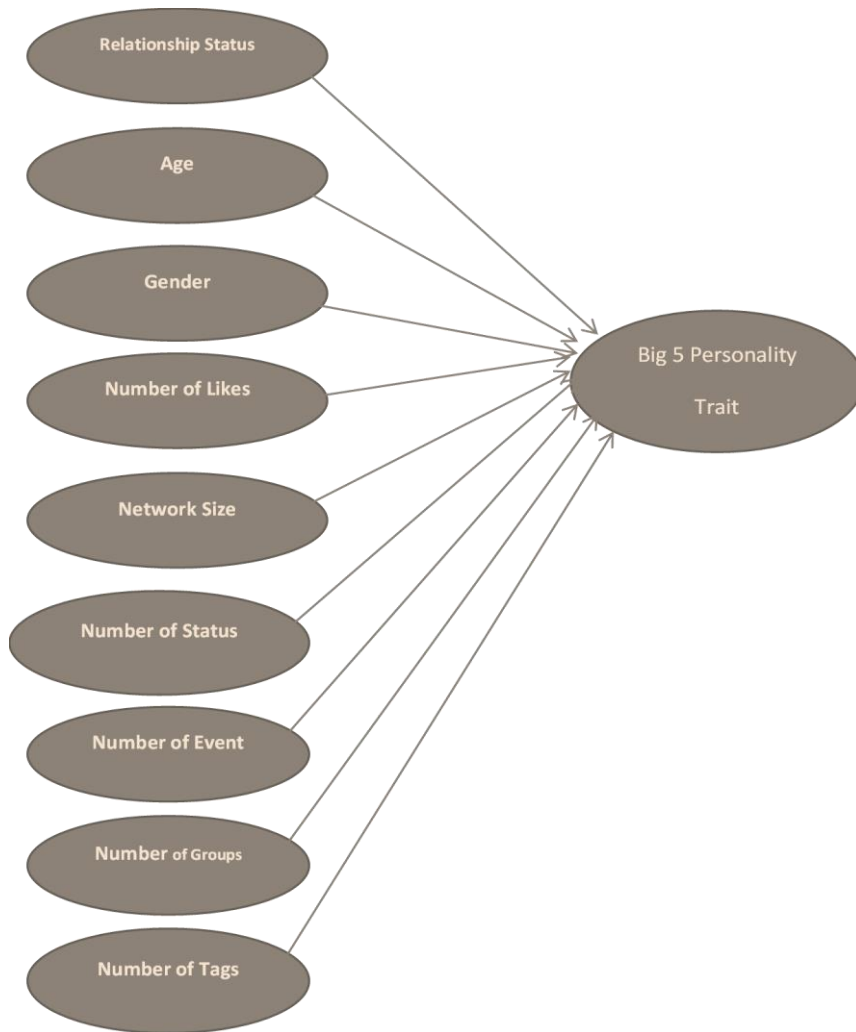


Figure 4.1: Predictive Model

4.2 Algorithm

The basic element of any good intelligent system is its algorithm, it's the algorithm that processes the data and uses it to produce knowledge. There are some steps that need to be taken when creating an algorithmic model. The first step and most crucial is the pre-processing of the data. This step is what prepares the data for the task to be carried out. Feeding data that has not be properly processed can greatly hamper the results of the model and can throw off the model completely. So before classifying or feeding the data into the ANN network first processing must be done. The next step is the actual processing itself which involves transforming the data received and then finally feeding it into the ANN for classification. Figure 4.2 shows the flow diagram of the model used in this study.

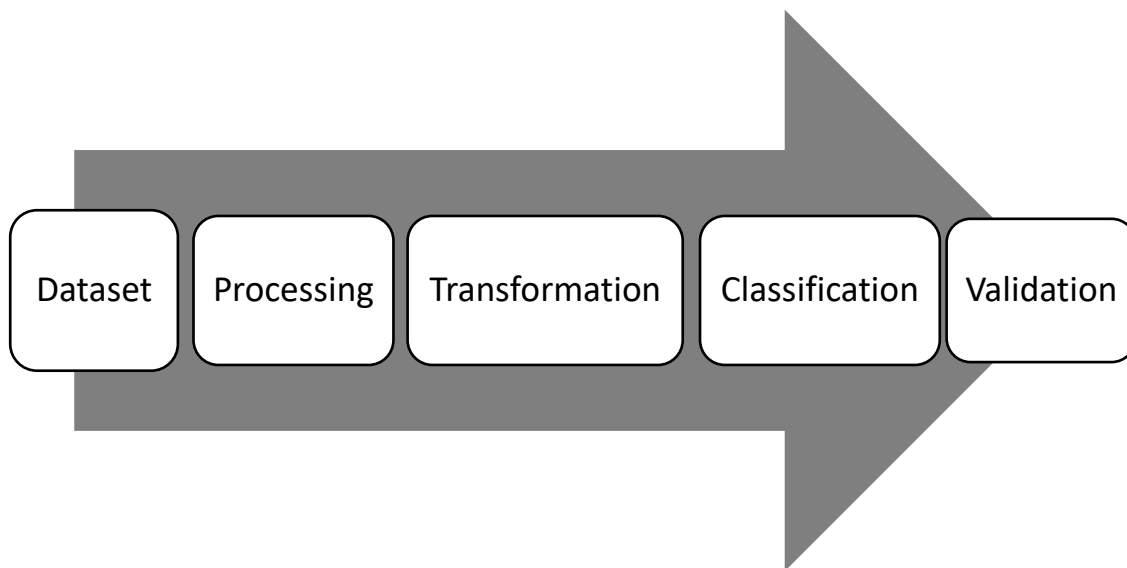


Figure 4.2: The Classification process of the BPNN Model

4.3 Data and Pre-processing

The Dataset used for this study was obtained from the database provided by the myPersonality project (Kosinski et al., 2015) which consist of Facebook data of over 4 million participants with given personality label which are based on the big 5 personality model. The myPersonality project was initiated by David Stillwell and Michal Kosinski. It is a Facebook application that

collects users Facebook information from their Facebook profile while taking privacy issues into consideration and also allows them take psychometric tests which calculates things such as satisfaction with life, big 5 personality, etc. The data retrieved from the application was processed, analyzed and then used to create the datasets. The data contain information of user's demographics, activities and friendship network size. During this study, the following the following datasets were downloaded.

- Big 5 model personality score: This table contained big 5 personality test scores taken by 3,137,694 Facebook users. It contained scores for the main big 5 traits Openness, Neuroticism, Agreeableness, Conscientiousness and Extraversion which results were scaled from 0 to 5.
- Facebook activity: This table contained a summary of the activities (tagging, posting, joining groups, tagging, etc) of 1,674,259 Facebook users. This table contained the number of likes, tags, updates, events, groups and friends.
- Demographics: In this table resides the basic attributes such as birthday, age, gender, relationship status, interest, time zone and network size of 4,282,857 unique Facebook users.

The dataset was downloaded from the myPersonailty database and needed to be merged together into one file, Microsoft SQL was used to merge the various database by their unique userid. The script used to merge the various database can be found in the appendix section.

The different database did not contain equal amount of participant, so in other to merge them, only common participant that could be found in all three Databases were merged, others that could not be found in the other databases were dropped. After merging the data, the dataset was left with 1337313 rows of unique participants with unique user IDs.

After the various database were successfully merged, missing values where considered. Missing data can greatly affect the results of a research because it could lead to biases, affect finding generalizability and result to a great loss of information (Dong & Peng, 2013). In other to ensure no missing values in the data, missing value analysis was done using IBM SPSS and it was observed that the merged file had a lot of missing values and before the database can be used for the study, missing values must be addressed. The some of the tables in the dataset had missing values above 50% so two methods had to be used to deal with the missing data, listwise deletion and replacing using series mean (Humphries, 2013). Another script was written in MSSQL in other to handle missing values and this script can be found in the appendix section.

Starting from the column with the highest missing, the script compares the values in the column with other columns and deletes that row if a missing value is found; this step was repeated across the columns until missing value was below 10%, eventually reducing the data to 7438 participants. After the missing values were reduced to 10% or less, the replace missing value option using series mean in SPSS was used to replace the remaining of the missing values. Now dataset only comprises of participants with no missing data, the data could now be further processed to be used in the neural network model. Out of the 7438, 3013(40.5%) were male and 4425(59.5%) were female.

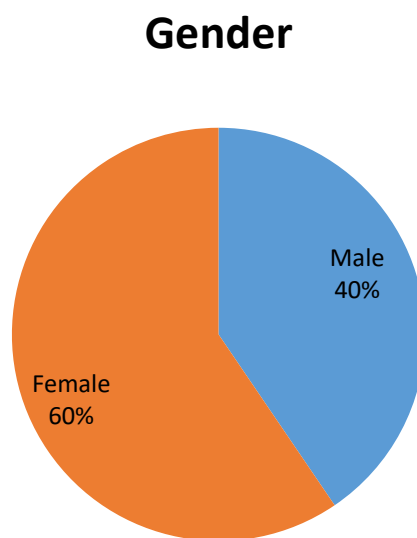


Figure 4.3: Distributions by Gender

The majority (57.4%) of the participants were in the age group of 18 to 25 years, followed by those (28.1%) within the age group of 26 – 40 years, then the 7.4% between 18years and below, the 6.4% within 40 – 60 years and also 0.7% those with 60 years and above. The Dataset also shows big 5 personality traits of the participants; 96% exhibited openness traits and 4% did not, 57% exhibit traits of neuroticism and 43% did not, 91% exhibited agreeableness traits and 9% did not, 87% exhibited conscientiousness traits but 13% did not and finally 88% exhibited extraversion traits and 12% did not.

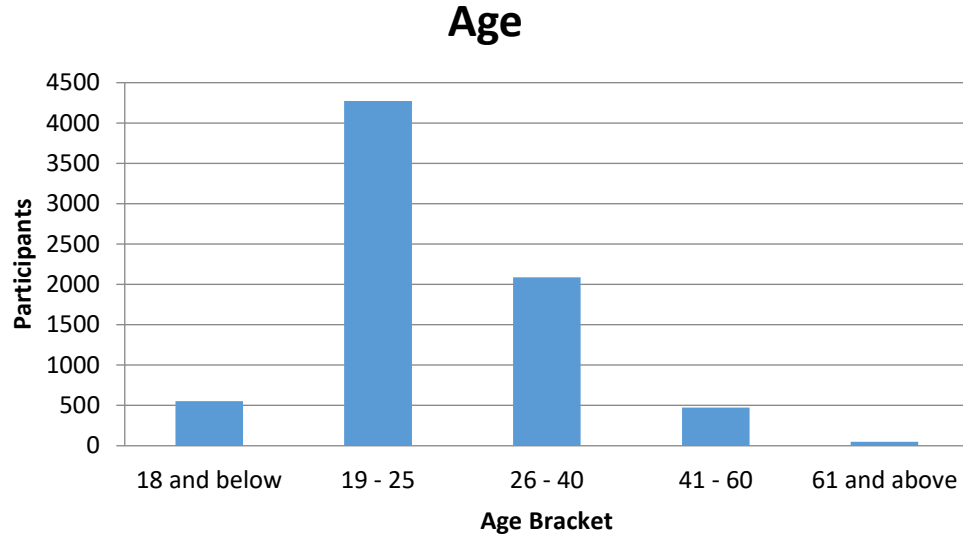


Figure 4.4: Distributions by Age

Table 4.1: Big 5 Personality Distribution

Value	OPE	NEU	AGR	CON	EXT
Yes	7181	4209	6789	6502	6567
No	257	3229	649	936	871

4.4 Transformation

When feeding the data into the neural network the data to be fed should be in tensors of floating point data(Chollet, 2017). The data also must not take widely different ranges because it could affect training. To ensure this is not the case, the best approach will be to normalize the data by transforming them into vectors of -1 to 1 or 0 to 1. Some of the data also might be scaled from 1 to 10. For example, a dataset can contain 10 countries and each country labelled 1 to 10 respectively or like in the case of this student relationship status was scaled from 1 to 10 for single, married, divorced, etc. Feeding the data this way could mislead the model to thinking 10 is greater than 1 meaning married is greater than single and that is not true. To handle this One-

Hot encoding is used. Currently this is the most common way to change tokens into vector(Chollet, 2017). It takes the total number of integer and then converts them into 0 and 1. If there are 10 variables, it takes the first variable and turns it into 1 and then turns the rest into 0's, it then takes the second integer, turns it into 1 and then turns the rest into 0's it continues this process for all the remaining integer. After the dataset has been successfully normalized and transformed it can be sent further for classification which outputs will return either 0 or 1.

4.5 Classification Architecture

Various methods have been used for classification in the past but in this study a multi label back propagation neural network technique was utilized for classification. ANN has been widely adopted into multi-label classification. The configuration and parameter in an ANN model needs to be selected efficiently so as to ensure appropriate generalization and efficient learning. What the model does is that through a feed forward process, updates itself by the back propagation update method and uses the supervised topology to enhance the model. This method is a multi-purpose learning algorithm, very effective and produces great results but it also costly in terms of learning requirement. With the help of a learning process and given hidden layer can simulate any function to any accuracy level (Helwan, Tantua, and Adeola, 2016). The hidden layer is the layer between the first and the last output layer. The data is taken from each of the input neurons through the synapses and multiplies it with a set of random weights. The summed weighted inputs are then passed through an activation function to the output layer. In this study two activation function was used ReLU and sigmoid activation function. The first to be used is the ReLU activation function

$$(A(x) = \max(0,1)) \quad (4.1)$$

ReLU is the most used activation function in the world today when sending signals from the first layer to the next layer before the output layer (SHARMA, 2017). Also since this is a multi-label classification problem and each label prediction probability needs to be predicted independently of the other class probabilities, sigmoid activate function is used as the second activation function.

$$\left(A = \frac{1}{1 - e^{-x}} \right) \quad (4.2)$$

Once this is done, the results are calculated against the actually expected targets to see how well they performed or how bad they performed, the error is what the model uses to make an update of the weights so as to go for the next round of iteration. The same method is carried out again and the weights are continually adjusted until the error comes to its lowest minima. The theory behind back propagation is error minimization and gradient descent. During the iteration process the error function of the model has to be a continuous derivable function so as to be able to find the least squared errors. To achieve this an activation function that can impose this is used some examples of this activation functions are logarithmic sigmoid functions or logistic regression loss (cross-entropy)

4.6 ANN Multi-Label Classification

In this study is an ANN task with multiple possible label samples that are not mutually exclusive meaning that a sample can have multiple label and not restricted to just one label, this is known as a multi-label classification problem. This problem is well tackled in ANN with a framework known as keras. In this study we have a problem with 5 different labels (openness, agreeableness, neuroticism, conscientiousness and extraversion) , therefore this study has n samples

$$X = \{x_1, \dots, x_n\} \quad (4.3)$$

and n number of labels

$$y = \{y_1, \dots, y_n\} \quad (4.4)$$

With $y_i \in \{1, 2, 3, 4, 5\}$ and $P(c_j | x_i)$ for the prediction probability. The next thing is to build a simple ANN with 5 output nodes and one output for each class. Designing the input and hidden layer is quite straight forward but designing the output layer for a multi-label and choosing what kind of layer it will be is quite important. Usually the softmax layer is the choice for multi class problem but this isn't really the best choice for a multi-label problem (Sterbak, 2017). In softmax when increasing score for one labels all others are lowered (probability distribution) which is not

a problem when predicting a single label per sample but in a multiple label prediction this is not good.

$$P(c_j | x_i) = \frac{\exp(z_j)}{\sum_{k=1}^5 \exp(z_k)} \quad (4.5)$$

What is needed is to decompose the multi-label classification task, for this, a sigmoid output layer is needed consisting of a sigmoid activation function and binary_crossentropy loss function. The labels will be improved individually and each label is independent of the other labels probability.

$$P(c_j | x_i) = \frac{1}{1 + \exp(-z_k)} \quad (4.6)$$

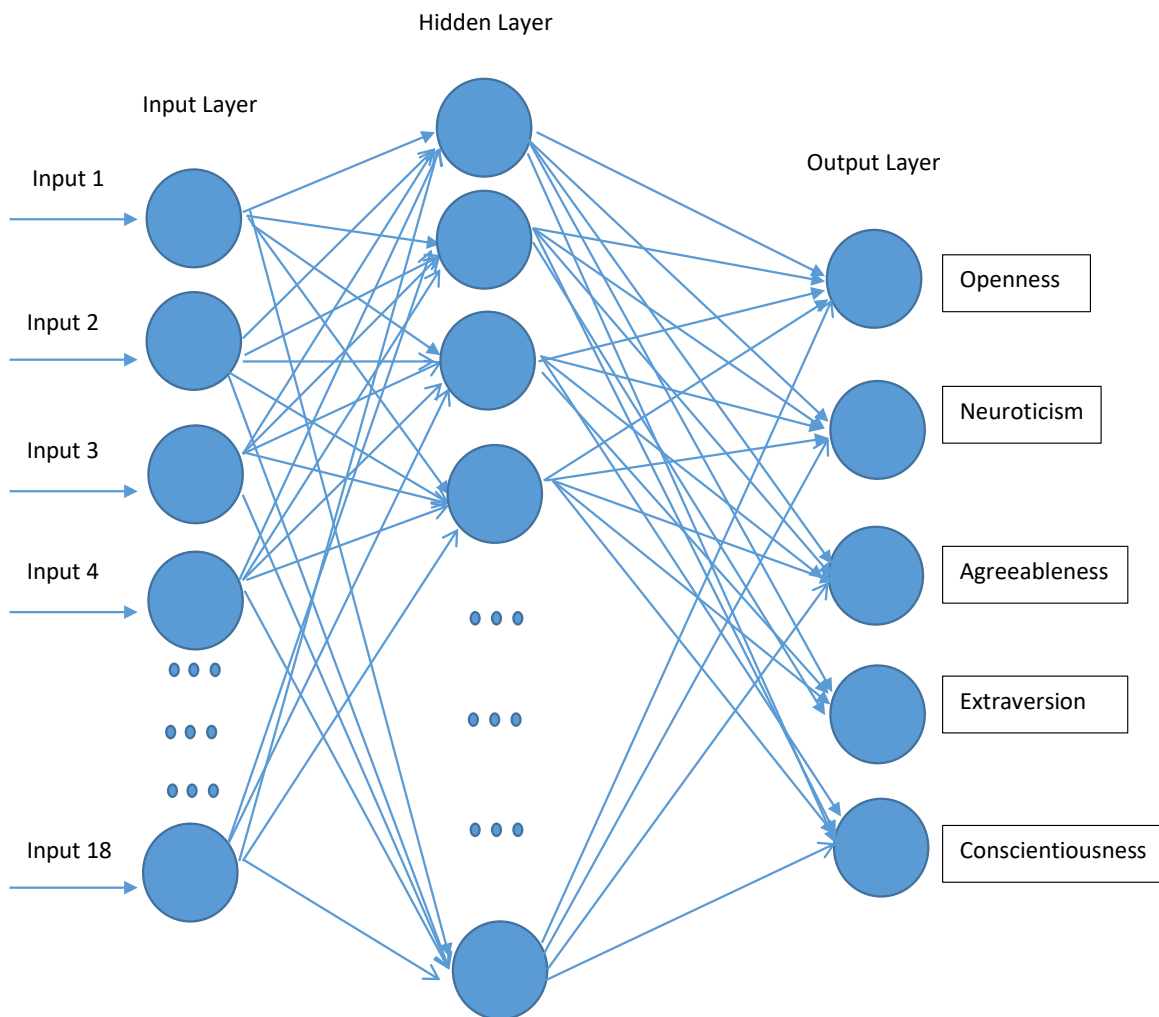


Figure 4.5: Neural Network model

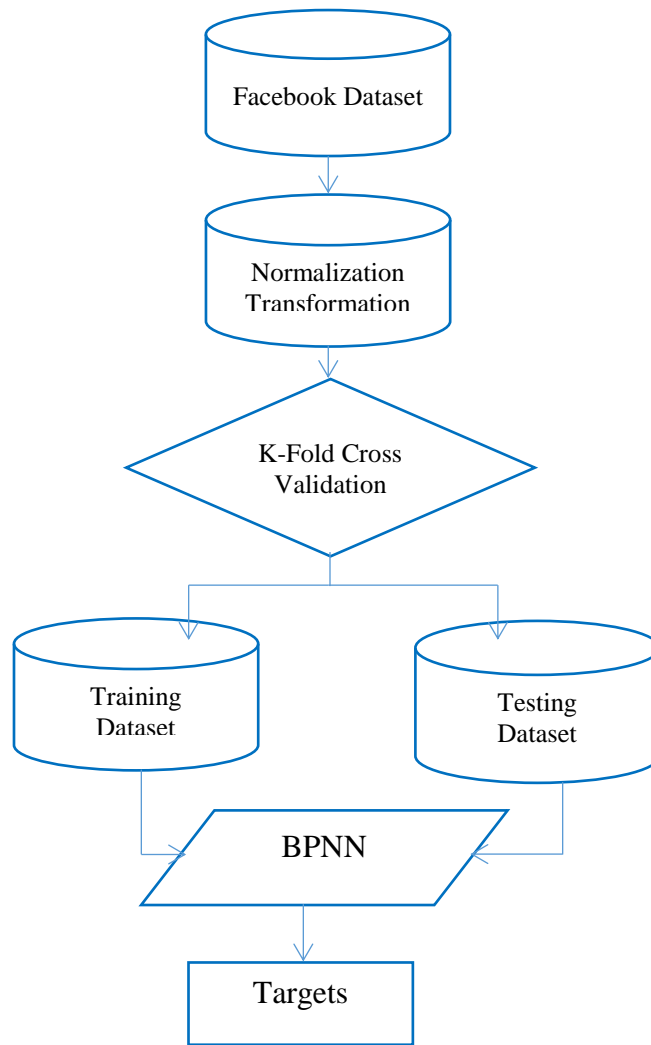


Figure 4.6: The Flow Diagram for study framework

4.7 Keras and Tensorflow

In this study the ANN model was built the user friendly Keras API which runs on Tensorflow. Tensorflow is an ANN learning package developed by google. By using the ANN architecture in tensorflow, building ANN has been made so easy just by applying basic implementations. In the tensorflow packages also includes tools which can easily be used for performance evaluation and helps provide means to manipulate the network so as to be able to learn better (Maxwell et al., 2017). The Data in Tensorflow are referred to as tensors. Various task such as normalization, vectorization or classification can be done on these tensors (Gerritsen, 2017). Today Tensorflow is gaining so much popularity because of its ease and simplicity, with tensorflow, ANN can be

created and trained without having to bother with the provision of gradient expression because they are computed automatically in tensorflow and works greatly on large dataset (Santolaya and Gavves, 2017).

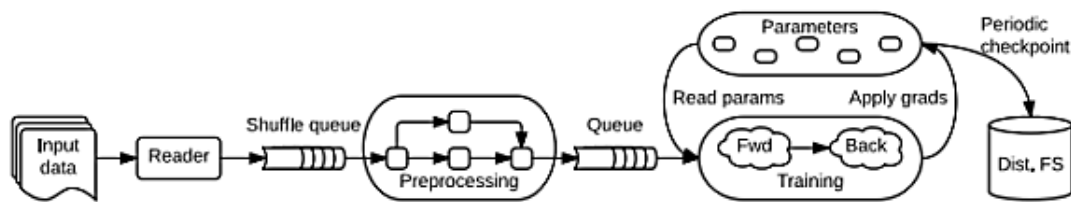


Figure 4.7: A Tensorflow Dataflow Diagram (Abadi et al., 2016)

To begin Due to the flexibility of Tensorflow, developers can experiment with different algorithms and optimization techniques with so much ease, also simplifying the implementation of new models. Tensorflow works on various multiple client languages but it has been prioritized on python.

Over the years the fame and usage of python has faced a massive exponential growth especially in the area of big data and machine learning and has become one of the most popular languages for data science (Chollet, 2017), as a matter of fact in 2017 it rose to the most wanted language in the year (Patel, 2017). The Python programming language is a language that's comprehensible by a wide range of people with a very understandable and readable syntax. Python also has libraries written for a number of tasks, especially matrix math operations making it very easy to use. Unlike other programming languages which will require a chunk of coding syntax just to execute a simple task, the vast libraries in python built to handle machine learning algorithm can get you off and running in no time. Also, it is a language with an active developer community (Harrington, 2012). Today carrying out an advanced deep learning research can be done easily with just a basic python scripting skill. This has been made possible by the development of two symbolic tensor manipulation frameworks for python; Theano and Tensorflow, greatly simplifying the process of new model implementation. The rise of user friendly libraries such as Keras has also made deep learning and ANN as easy as LEGO manipulation (Chollet, 2017)

4.8 Training, Testing and Validation

When working with classification techniques the dataset is separated into training, testing and validation set that enables you to monitor accuracy of the model on data.

Training is the learning process of feeding quality set of training example data known as training set repeatedly to the network. This training is important in other to enable the network build up its classification mechanism. The quality of a network is calculated by first training it then validating it with the validation data, when it looks good it is then finally tested with the testing data set, which is the hold out data kept aside after the network has been trained.

The validation and testing set is created from the original set by setting it apart during training. There is no standard way on how to decide how much to allocate for training, testing and validation but some approaches has gained popularity. For example, in large dataset 60% is allocated to training, 20% to testing and 20% for validation. However, in a small dataset some researches prefer not to split dataset this way but use cross validation technique (Helwan et al., 2017). In cross validation before the training begins, a portion of the dataset is kept aside, after training is done, the portion that was kept aside is used to test the model.

There are two common approaches used to evaluate the performance of a classification model that are K-fold and leave one out validation (Wong, 2015).

The K-fold cross validation is a very popular method to estimate the performance of a classification algorithm. In this procedure the dataset is randomly broken down and split into k-partitions with equal sizes and the holdout technique is iterated K (e.g 5, 10, etc) times. Identical K models are instantiated forming training sets on the k-1 partitions while some are held back for evaluation. In regards to the number of iterations, the K validations scores average then becomes the validation score for the model. The advantage of K- fold validation is that all the point in the K-fold will all appear for the training and for the testing so the order in which they are divided is of no importance. The strength and variant of this approach is the random division of dataset to training and testing in K for various times.

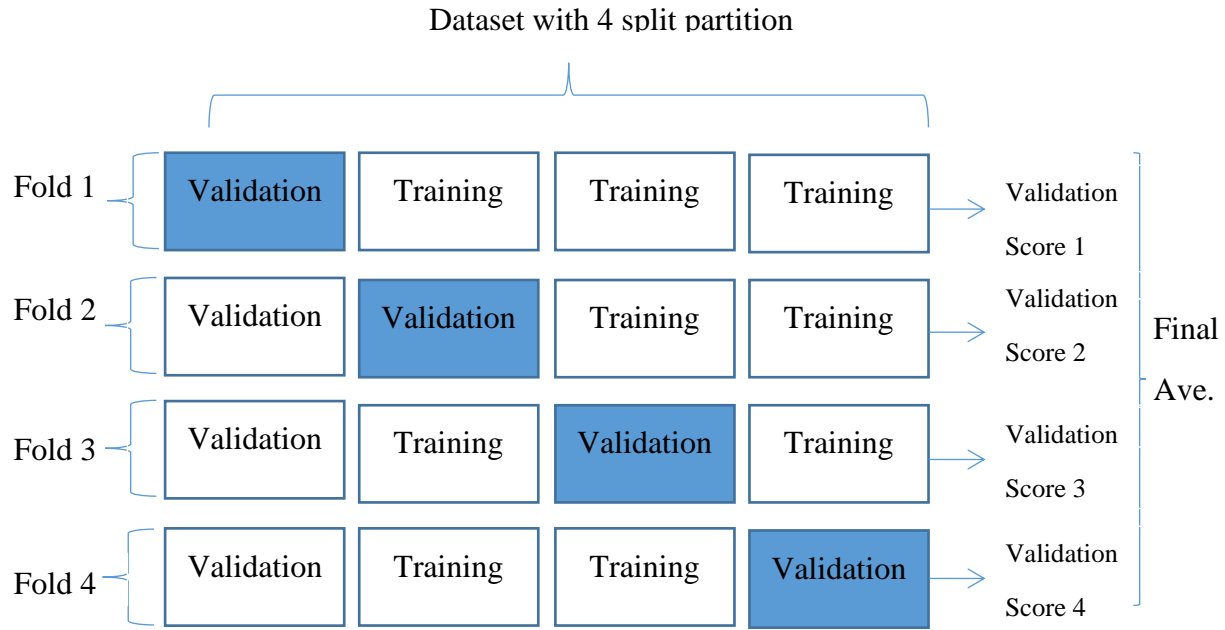


Figure 4.9: K-Fold Cross Validation

Another type of cross validation many studies adopt is the leave one out cross validation. Most studies use this approach when the instances in a data set is small. The instances that divided for training and testing are not randomly selected the estimate point of accuracy for a particular data is set at a constant value (Wong, 2015). In this approach, K is equivalent to N which is the number points in a dataset. The function carries out the training on the entire dataset but leaves out point, which it then conducts the prediction on just that point.

In this study the K-fold cross validation was used for this study to conduct the evaluation of the model

4.9 Visualization

In other to properly make decisions and better understand the learning process, visualization is very important especially when dealing with large dataset. This is because it enables a better understanding of the classification result (Chae et al, 2017). The Matplotlib for python, which is a 2D plotting library, was used to enable the training process and pattern of the model to be properly visualized.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Experimental Setup

Due to the high availability of the machine libraries and ability to carry out amazing data visualization, Python was used to implement the data processing and modelling. The neural network was built taking advantage of the rich library in the Keras API running on Tensorflow. In the building of the neural network it is important to identify the number of hidden layers and hidden neurons to be used in the network. Depending on the data it is safe to start with a few hidden layers preferably one fully connected hidden layer (Chollet, 2017). The model was setup with a fully connected hidden layer between the input layer and the output layer. The ReLU was used to act as the non-saturating activation function between the first layer and the hidden layer while the sigmoid activation function was used as the final output activation function. When deciding for the number of hidden neurons it could be somewhat complicated to select the best number of hidden neurons perfect for the task without examining several models. Inadequate or too much hidden neurons could lead to over fitting or underfitting. In neural network a lot of try error is usually done to ascertain the best parameters for the network. There is some rule of thumbs when deciding the number of hidden neurons (Heaton, 2008)

- It should be less than two times the input layer size
- $\frac{2}{3}$ of the sum of input and output neurons

In this study different number of hidden neurons were tried and then increased to ascertain the neurons with best performance.

The data was imported using the panda's data frame because of the massive functionality the panda's data frame gives to work on data. After data has been successfully imported the next step is to create a matrix of the features and target variable, this enables the network to identify the input and the output file, 9 input variables and 5 output variables. As discussed in the previous section features such as relationship status were coded from 1 to 10 but leaving it this way would mean 10 is higher than 1, meaning single is higher than divorced which is not the

case. To handle these, dummy variables were created using a function from the SciKitLearn library in python known as OneHotEncoder (Pedregosa et al., 2011). A dummy variable is a dichotomous variable that has been coded to represent a variable with a higher level of measurement. The dummy variables are split into n number of variables in this case n will be 10 because the relationship status is coded from 1 to 10. Each variable represents 1 and the other which isn't the variable is represented as 0. Figure 5.1 shows a sample of the data set before one hot encoding and figure 5.2 shows a sample of the table after one hot encoding.

133	5	1	1	30	236	79	16	90	4	3.75	2.5	4.75	4.5
104	2	2	0	26	521	14	53	129	3.8	1.8	3.7	3.3	3.8
551	1	8	1	23	100	278	23	206	3.4	2.3	3.1	4.3	2.6
212	2	66	0	21	288	15	28	206	3.95	3.3	1.8	2.65	3.6
70	2	6	0	25	395	67	9	64	4.9	3	4	1.8	3.9
197	2	7	1	19	242	332	4	269	4.5	3.33	4.75	4.5	3.5
303	1	1	0	22	256	45	3	141	4	2.75	4	4.5	4
48	3	1	1	29	462	211	44	165	3.25	1.67	4.25	4	4
261	2	3	1	18	134	850	73	302	4.65	2.95	3.5	3.6	3.65
144	3	120	1	27	247	107	62	436	4.25	2	4.13	4.38	4.25
485	1	5	0	21	25	123	44	134	4.4	3.1	3.7	3.9	3.35
601	2	54	1	19	785	830	297	166	3.9	3.8	2.95	3.45	4.5
47	1	15	1	21	816	14	44	10	3.5	2.75	3.55	3.45	3.84
377	3	11	1	20	208	321	151	90	3.8	1.79	3.95	3.89	4.25
17	1	14	1	23	347	5	36	364	4.38	2.5	3	2.13	3.5
480	2	2	1	24	742	103	130	679	4.5	3.25	3.5	4.5	4.75
14	1	32	0	23	315	77	26	1305	3.8	3.5	1.55	2.45	3.8
473	3	2	1	41	52	71	14	179	3.4	4.05	3.45	2.4	2.05
195	1	3	1	21	743	174	95	255	3.29	3.07	3.57	3.86	2.93
638	6	1	0	21	101	252	14	68	4.9	3	2.9	3.4	4.55
652	2	6	0	26	97	35	44	5	3.25	1.5	2.25	4.25	4.25
871	2	2	1	24	406	138	74	145	4.35	1.95	3.37	4.65	4.6
151	3	9	0	30	237	24	13	110	4.4	1.2	4.5	3.35	3.58
601	1	7	0	22	263	5	15	341	2.25	3.75	3	3	3
121	1	1	0	22	211	4	9	122	3.5	2.5	4.25	4	3

Figure 5.1: Before one hot encoding

0	0	0	1	0	0	0	0	0	0	133	1	1	30	236	79	16	
1	0	0	0	0	0	0	0	0	0	104	2	0	26	521	14	53	1
0	0	0	0	0	0	0	0	0	0	551	8	1	23	100	278	23	2
1	0	0	0	0	0	0	0	0	0	212	66	0	21	288	15	28	2
1	0	0	0	0	0	0	0	0	0	70	6	0	25	395	67	9	
1	0	0	0	0	0	0	0	0	0	197	7	1	19	242	332	4	2
0	0	0	0	0	0	0	0	0	0	303	1	0	22	256	45	3	1
0	1	0	0	0	0	0	0	0	0	48	1	1	29	462	211	44	1
1	0	0	0	0	0	0	0	0	0	261	3	1	18	134	850	73	3
0	1	0	0	0	0	0	0	0	0	144	120	1	27	247	107	62	4
0	0	0	0	0	0	0	0	0	0	485	5	0	21	25	123	44	1
1	0	0	0	0	0	0	0	0	0	601	54	1	19	785	830	297	1
0	0	0	0	0	0	0	0	0	0	47	15	1	21	816	14	44	
0	1	0	0	0	0	0	0	0	0	377	11	1	20	208	321	151	
0	0	0	0	0	0	0	0	0	0	17	14	1	23	347	5	36	3
1	0	0	0	0	0	0	0	0	0	480	2	1	24	742	103	130	6
0	0	0	0	0	0	0	0	0	0	14	32	0	23	315	77	26	13
0	1	0	0	0	0	0	0	0	0	473	2	1	41	52	71	14	1
0	0	0	0	0	0	0	0	0	0	195	3	1	21	743	174	95	2
0	0	0	0	1	0	0	0	0	0	638	1	0	21	101	252	14	
1	0	0	0	0	0	0	0	0	0	652	6	0	26	97	35	44	
1	0	0	0	0	0	0	0	0	0	871	2	1	24	406	138	74	1
0	1	0	0	0	0	0	0	0	0	151	9	0	30	237	24	13	1
0	0	0	0	0	0	0	0	0	0	601	7	0	22	263	5	15	3

Figure 5.2: after OneHotEncoding

After one hot encoding the transformation and normalization were also done using the SciKit-Learn library. The dataset is transformed into vectorised form from 0 to 1 so as to enable the network better understand the data for classification. The output target dataset was also transformed using 0.5 as its threshold. The SciKit-Learn is also used to split the data set into training and test samples.

0	0	0	1	0	0	0	0	0	0	0.054388	0	1	0.254545	0.058912	0.017322	0.025424	0.0
1	0	0	0	0	0	0	0	0	0	0.042439	0.000465	0	0.218182	0.130358	0.002887	0.088136	0.0
0	0	0	0	0	0	0	0	0	0	0.226617	0.003253	1	0.190909	0.024818	0.061515	0.037288	0.0
1	0	0	0	0	0	0	0	0	0	0.086939	0.030204	0	0.172727	0.071948	0.003109	0.045763	0.0
1	0	0	0	0	0	0	0	0	0	0.02843	0.002323	0	0.209091	0.098772	0.014657	0.013559	0.0
1	0	0	0	0	0	0	0	0	0	0.080758	0.002788	1	0.154545	0.060416	0.073507	0.005085	0.1
0	0	0	0	0	0	0	0	0	0	0.124433	0	0	0.181818	0.063926	0.009771	0.00339	0
0	1	0	0	0	0	0	0	0	0	0.019365	0	1	0.245455	0.115568	0.046636	0.072881	0.0
1	0	0	0	0	0	0	0	0	0	0.107128	0.000929	1	0.145455	0.033342	0.188541	0.122034	0.1
0	1	0	0	0	0	0	0	0	0	0.05892	0.055297	1	0.227273	0.06167	0.02354	0.10339	0.1
0	0	0	0	0	0	0	0	0	0	0.199423	0.001859	0	0.172727	0.006017	0.027093	0.072881	0.0
1	0	0	0	0	0	0	0	0	0	0.247219	0.024628	1	0.154545	0.19654	0.184099	0.501695	0.0
0	0	0	0	0	0	0	0	0	0	0.018953	0.006506	1	0.172727	0.204312	0.002887	0.072881	0.0
0	1	0	0	0	0	0	0	0	0	0.154924	0.004647	1	0.163636	0.051893	0.071064	0.254237	0.0
0	0	0	0	0	0	0	0	0	0	0.006593	0.006041	1	0.190909	0.086739	0.000888	0.059322	0.0
1	0	0	0	0	0	0	0	0	0	0.197363	0.000465	1	0.2	0.185761	0.022652	0.218644	0.3
0	0	0	0	0	0	0	0	0	0	0.005356	0.014405	0	0.190909	0.078716	0.016878	0.042373	0
0	1	0	0	0	0	0	0	0	0	0.194479	0.000465	1	0.354545	0.012785	0.015545	0.022034	0.0
0	0	0	0	0	0	0	0	0	0	0.079934	0.000929	1	0.172727	0.186012	0.038419	0.159322	0.0
0	0	0	0	1	0	0	0	0	0	0.262464	0	0	0.172727	0.025069	0.055741	0.022034	0.0
1	0	0	0	0	0	0	0	0	0	0.268232	0.002323	0	0.218182	0.024066	0.007551	0.072881	0.0
1	0	0	0	0	0	0	0	0	0	0.358467	0.000465	1	0.2	0.101529	0.030424	0.123729	0.0
0	1	0	0	0	0	0	0	0	0	0.061805	0.003717	0	0.254545	0.059163	0.005108	0.020339	0.0
0	0	0	0	0	0	0	0	0	0	0.247219	0.002788	0	0.181818	0.065681	0.000888	0.023729	0.1

Figure 5.3: Sample of input data after transformation

ope	neu	agr	con	ext	
1	1	0	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	1	0	1	1	1
1	1	1	0	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	0	1	1	1	1
1	1	1	1	1	1
1	0	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	0	1	1	1	1
1	0	1	0	1	1
1	1	1	1	1	1
1	1	0	0	1	1
1	1	1	0	0	1
1	1	1	1	1	1
1	1	1	1	1	1
1	0	0	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
0	1	1	1	1	1

Figure 5.4: Sample of the Output Layer after Transformation

5.2 Training and Testing

The training and testing for this study was broken down into two different schemes, the first scheme involved manually splitting the dataset into 2 sets, test set and training set. The first scheme was also split into different parts also so as to experiment on different parameters. In the second scheme the training and testing was done with the K-Fold cross validation. This also has two different tuning parameters so as to examine which will produce the best results

5.2.1 Training

In the first scheme, the first part was made up of 75% training data and 25% testing data that the model has not seen before while the second part was made up of 67% training data and 33% testing data.

In the Second scheme personality classification model was trained using K-10 fold cross validation and K-5 Fold cross validation. On the K-10 training examines 6695 cases which is about 90% of the data and tested with 743 cases consisting of about 10% of the data, this was automatically split by the K-Fold cross validation 10 times and at each time, the folds were selected randomly. Also in the K-5 Fold training examines 5951 cases which is about 80% of the data and tested with 1487 cases consisting of about 20% of the data, this was automatically split by the K-Fold cross validation 5 times and at each time, the folds were selected randomly. In the final process of the BPNN model, the input neuron was made up of 18 input neurons to accommodate for the 18 features. For the 5 personality classes; openness, agreeableness, neuroticism, extraversion and conscientiousness, the output layer was made up of 5 output neurons. For the sake of optimization different parameters were set, such learning rate, hidden neurons and splits, the results were compared with each of them yielding different results. The maximum epochs was set to 1000 but the network was told to end training if network loss performance keeps declining and doesn't improve after 10 epochs. All the training and computation was carried out using a 3.30 GHz PC with 8GB RAM, intel core i5 and windows 10 OS. Table 5.1 shows the parameters for the different networks.

Table 5.1: Back propagation neural network training parameter

Parameters	Scheme 1(No K Fold Cross Validation)		Scheme 2(KFold Cross Validation)	
	Test 1	Test 2	Test 1	Test 2
Number of training samples	(75:25)	(67:33)	10 Fold (90:10)	5 Fold (80:20)
Number of Hidden Neurons	15	30	15	30
Learning rate	0.001	0.0001	0.001	0.0001
Maximum Epochs	1000	1000	1000	1000
Training Time (Secs)	33.79	21	188	82
loss	0.3654	0.3675	0.3631	0.3650
Training Method	Adam	Adam	Adam	Adam
Hidden Layer Activation Function	ReLU	ReLU	ReLU	ReLU
Hidden Layer Activation Function	Sigmoid	Sigmoid	Sigmoid	Sigmoid
Dropout percentage	0.3	0.3	0.3	0.3

From table 5. 1 it can be seen that Scheme 2 Test 1 has the lowest loss of 0.3631 concluded in 188 seconds after 10 fold cross validation. Unlike accuracy, loss is not a percentage, it is the summation of the errors made for each example in training and validation sets. In this case of neural networks, the loss is the negative log-likelihood (binary cross entropy). The learning curve showing the convergence of the network plots the loss against the epochs and the accuracy against the epochs is shown in figure 5.5, 5.6, 5.7, 5.8 representing scheme 1 test 1, scheme 1 test 2, scheme 2 test 1 and scheme 2 test 2 respectively.

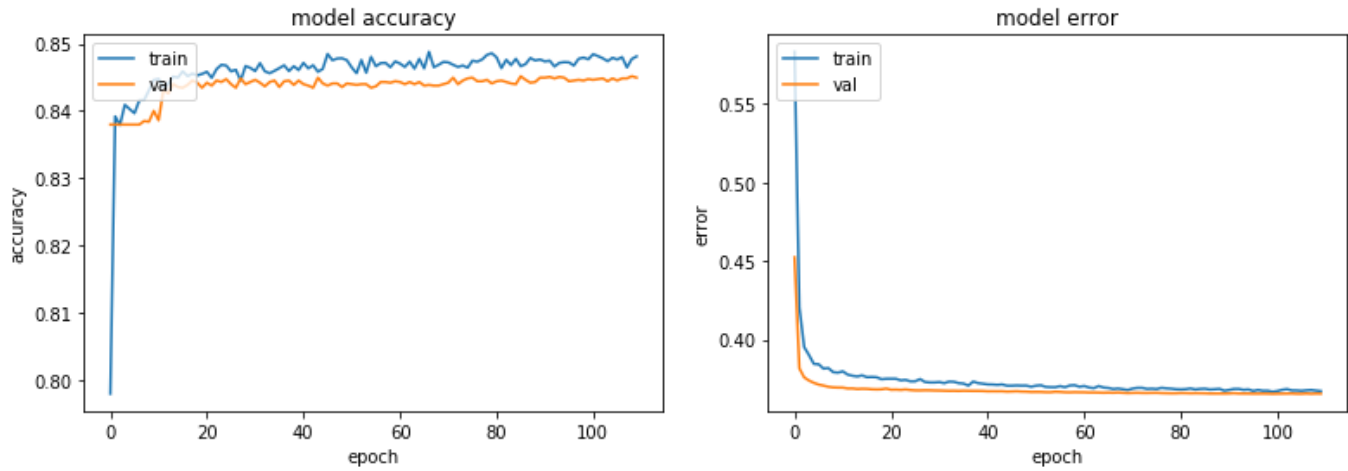


Figure 5.5: Accuracy and Loss for Scheme 1 Test 1(75:25 Split)

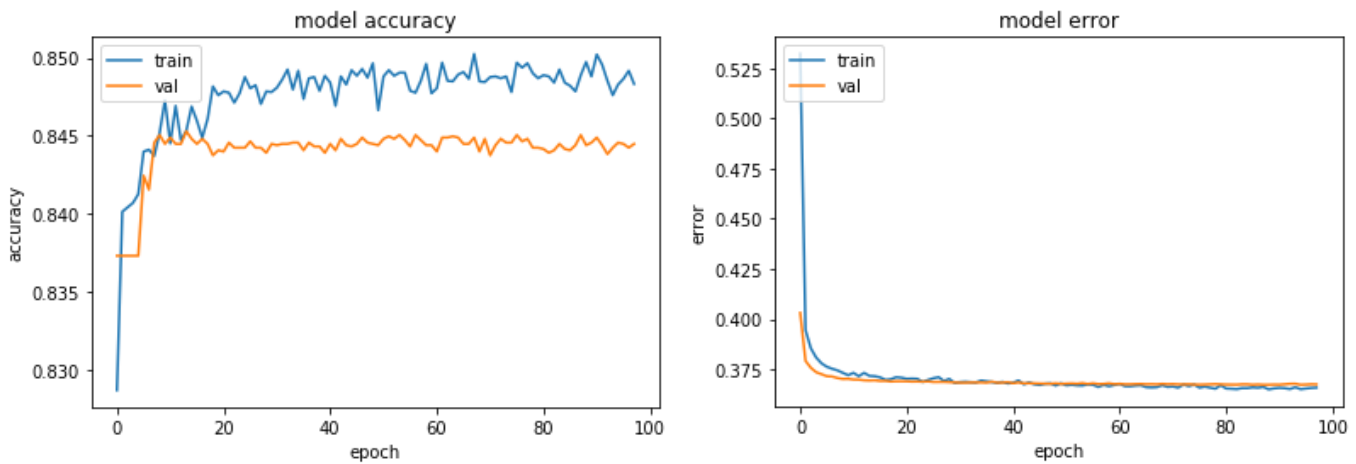


Figure 5.6: Accuracy and Loss for Scheme 1 Test 2(67:33 split)

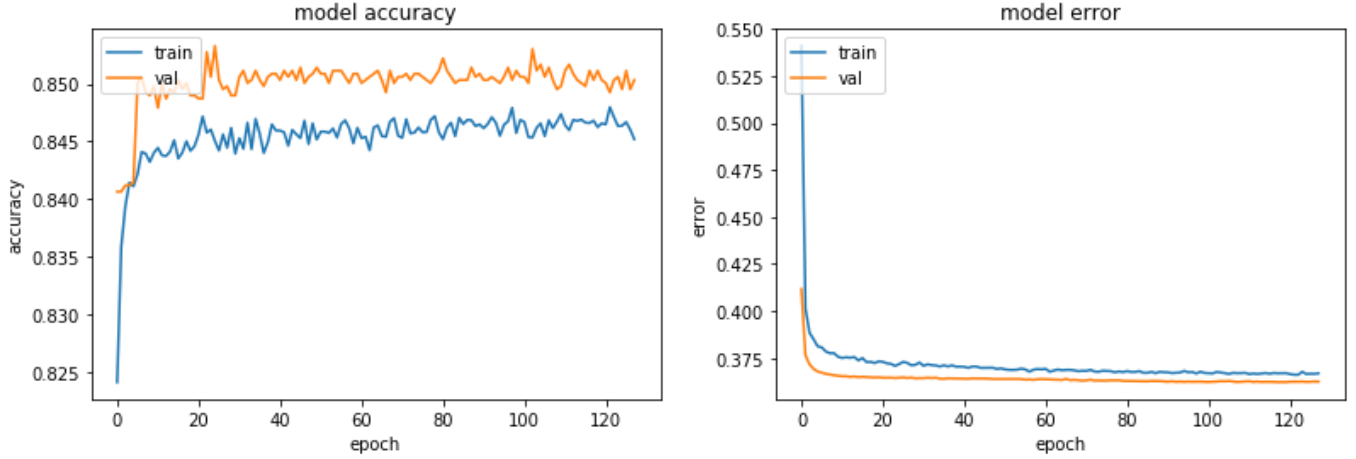


Figure 5.7: Accuracy and Loss for Scheme 2 Test 1(K-10 Fold)

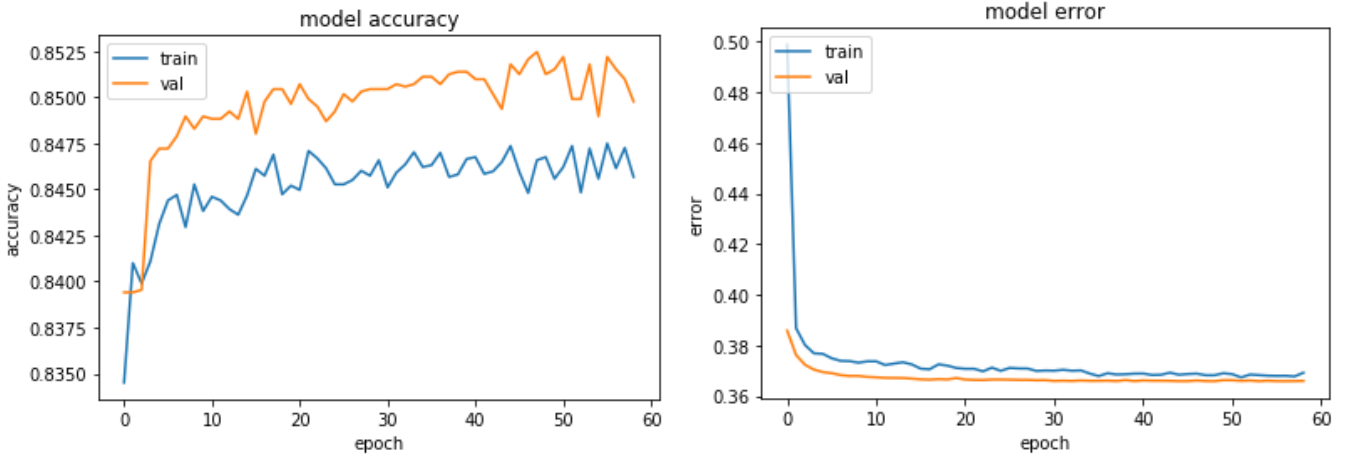


Figure 5.8: Accuracy and Loss for Scheme 2 Test 2(K-5 Fold)

5.2.2 Testing

Scheme 1 test 1 and test 2 were both tested on 25% and 33% of the unseen data, while the networks for scheme 2, test 1 and test 2 were tested and validated using a 10-fold Cross validation and 5-fold cross validation respectively. The data was trained on 6695 samples and tested with 743 unseen data, also trained again on 5951 samples and tested on 1487 unseen data. The model performs quite well with the test data and shows good generalizability.

This is to say that the model gives compelling generalization abilities when presented with new users Facebook Activity Data. Table 5.2 shows the prediction accuracy results and the hamming loss for all the trained networks. The hamming loss is part of the metrics used for evaluation in this study. It is used to compute accuracy through the equilibrium contrast between the target data and predicted data.

The hamming loss is the fraction of labels that are incorrectly predicted. The best hamming loss for the model is 14.96% which therefore implies that in more than 85% of the time, the model can correctly classify an individual based on their Facebook activity

Table 5.2: Back propagation neural network training and testing results

Network Parameter	Scheme 1(No K Fold Cross Validation)		K-10 Fold Cross Validation	
	Test 1	Test 2	Test 1	Test 2
Number of training samples	(75:25)	(67:33)	10 fold (90:10)	5 Fold (80:10)
Correctly Classified Training Samples	(4720/5578)	(4221/4983)	(5678/6695)	(5046/5951)
Recognition Rate on Training	84.62%	84.71%	84.81%	84.79%
Number of Test Samples	1860	2455	743	1487
Correctly Classified Test Samples	1571	2073	630	1265
Recognition Rate on Testing	84.47%	84.46%	85.01%	85.04%
Hamming Loss	15.53	15.54	14.99	14.96
Overall Recognition Rate	84.54%	84.58%	84.84%	84.92%

5.2.3 The Results Discussion

Different studies have been carried out on the subject of using various means to infer personality. Some significant contributing studies have used real life activities pertaining to locations and speech to infer personality (Mairesse et al., 2007). Another significant contribution by (Golbeck, Robles, & Turner, 2011) took a different approach and focuses on using social media to infer personality and discovered a tight relationship between a user's profile and personality. Some other studies by Bachrach et al. (2012) and Kosinski et al. (2013) shows some major inferences between demographic attributes and activities to a person's personality.

These studies provided motivations to carry out this study. The influence of social media and especially Facebook in the society is rapidly on the increase. Facebook has become an integral part of our lives, businesses, government and many more. It is important to understand what extent of inference Facebook has to a person's personality so that this can be used to better improve lives, safe guard lives and better society. In this study 2 inference models from three difference studies were combine, this being the features highlighted by Bachrach et al. (2012) and Sumner et al. (2011) with the features highlighted by Kosinski et al. (2013). The developed framework shows its capacity to predict the big 5 personality traits; Openness. Agreeableness, Conscientiousness, Extraversion and Neuroticism from a user's Facebook data. As shown in this study this ANN model shows encouraging results, in that in 85% percent of the time the network will correctly classify a Facebook user based on just their activities. Upon trying different methods the best classification result derived from the model had a prediction accuracy of 85.04% although their differences where not so distinct. Comparing the results of this study with some other studies such Tandra et al. (2017) and Lima & de Castro (2014) networks, this model performs better. In the study by Tandra et al. (2017) which used linguistic features on Facebook data to infer personality was able to hit 70% accuracy on their neural network model, while in the study by Lima & de Castro (2014), a semi supervised learning approach was used and the outputs where broken down and analyzed separately into five different binary outputs, this method gave a 75% prediction accuracy. In this study the proposed models analyzing just Facebook activities and demographics alone was able to perform better with a prediction accuracy of 85%. This shows how the ANN model can be used to learn accurately and faster during the training phase with if given more data however the generalizability is weakened in different scenarios which maybe a result of the data or the parameters used during the training. Given suitable pre-processing and adequate amount of dataset, this present study evinces the viability of ANN models for personality classification and it also shows the usefulness.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

The study ends with a concluding remark and further future recommendation.

6.1 Conclusion

The purpose of this study was to explore the performance of ANN in classifying and predicting the big five personality based on the data derived from a user's Facebook data. This study proffers an apt system classification model for big five-personality prediction that could accurately infer an individual's personality based on only their Facebook data with a prediction accuracy of 85.04%. The observations showed that ANN with proper parameter tuning could perform well in accuracy on complex multi-label task such as personality classification when trained and tested with new data. With the rapid growth in demand amongst various companies in better understanding their clients, this has increased the demand for online tools that can help better under the personality of the consumers

One of the limitations of this study is that a huge amount of data was lost during data pre-processing but more data can be thrown to the model to try to improve training. To improve the model training quality there is a need for exponential more data, so much data can be gotten from an individual's social media account. Another limitation to this study is that accuracy was not verified with other methods such as partial least squares and other machine learning methods. This similar study should be carried out on the same participants other accounts so as to better compare results and improve prediction. Finally, more studies should be carried out in this area of utilizing neural network to better understand and predict personality so as to understand ways to make people's lives better; With the prediction accuracy improved more, this model can be implemented in Facebook, users will no longer need to fill long personal forms to be able to determine their personality type, the personality type can be determined on Facebook just from user's activities without having to fill any forms. Users can be able to make results public and share on their wall. In business, based on the requirement of company, organizations can be able to predict the personality of workers to see how they can better improve their service. Advertisers can better know how to target their audience, for instance, advertisers can target

people with openness personality when advertising new products or target people with neurotic personality when advertising security products. The data that can be retrieved from Facebook data is very rich, further studies can be carried out in combining ANN with big five personality traits to analyze Facebook data to predict depression and suicidal tendencies.

REFERENCES

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Brain, G. (2016). TensorFlow: A System for Large-Scale Machine Learning. *In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16) (pp. 265–283)*. Savannah,GA.
- Adali, S., and Golbeck, J. (2012). Predicting Personality with Social Behavior. *In 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 302–309)*.
- Adali, S., Sisenda, F., and Magdon-Ismail, M. (2012). Actions speak as loud as words. *In Proceedings of the 21st international conference on World Wide Web - WWW '12 (p. 689)*. New York, New York, USA: ACM Press.
- Agrawal, A. (2017). Loss Functions and Optimization Algorithms. Demystified. Retrieved May 4, 2018, from <https://medium.com/data-science-group-iitr/loss-functions-and-optimization-algorithms-demystified-bb92daff331c>
- Akshat, D. (2016). Application of convolutional neural network models to personality prediction from social media images and citation prediction for academic papers. UNIVERSITY OF CALIFORNIA,SAN DIEGO.
- Al-Shihi, H., Sharma, S. K., & Sarrab, M. (2018). Neural network approach to predict mobile learning acceptance. *Education and Information Technologies*.
- Anderson, M., Antenucci, D., Bittorf, V., Burgess, M., Cafarella, M., Kumar, A., ... Zhang, C. (2013). Brainwash: A Data System for Feature Engineering. *In Conference on Innovative Data Systems Research (CIDR)*.
- Asur, S., and Huberman, B. A. (2010). Predicting the Future with Social Media. *In 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (pp. 492–499)*. IEEE.
- Ateş, U. (2014). Inference of Personality Using Social Media Profiles, (June).
- Auchard, E., and Ingram, D. (2018). Cambridge Analytica CEO claims influence on U.S.

- election, Facebook questioned. Retrieved April 7, 2018, from <https://www.reuters.com/article/us-facebook-cambridge-analytica/cambridge-analytica-ceo-claims-influence-on-u-s-election-facebook-questioned-idUSKBN1GW1SG>
- Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., and Stillwell, D. (2012). Personality and patterns of Facebook usage. In *Proceedings of the 3rd Annual ACM Web Science Conference on - WebSci '12* (pp. 24–32). New York, New York, USA: ACM Press.
- Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., and Gosling, S. D. (2010). Facebook Profiles Reflect Actual Personality, Not Self-Idealization. *Psychological Science*, 21(3), 372–374.
- Bataineh, M. H., Abdel-Malek, K., and Marler, T. (2012). *Artificial neural network for studying human performance*. University of Iowa.
- Binh, H. T., and Duy, B. T. (2017). Predicting students’ performance based on learning style by using artificial neural networks. In *2017 9th International Conference on Knowledge and Systems Engineering (KSE)* (pp. 48–53).
- Bronstein, A. (2017). Train/Test Split and Cross Validation in Python – Towards Data Science. Retrieved May 4, 2018, from <https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6>
- Champa, H., and AnandaKumar, K. (2010). Artificial Neural Network for Human Behavior Prediction through Handwriting Analysis. *International Journal of Computer Applications*, 2(2), 975–8887.
- Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., ... Zhang, Z. (2015). MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems. *arXiv preprint arXiv:1512.01274*.
- Chollet, F. (2017). *Deep Learning with Python*. Manning Publications Co..
- Chowdhury, A. P. (2017). AI in Dating Apps: Machine Learning comes to the rescue of dating apps. Retrieved April 17, 2018, from <https://analyticsindiamag.com/ai-dating-apps-machine-learning-comes-rescue-dating-apps/>

- Ciodaro, T., Deva, D., de Seixas, J. M., and Damazio, D. (2012). Online particle detection with Neural Networks based on topological calorimetry information. *Journal of Physics: Conference Series*, 368(1), 12030.
- Corani, G., and Scanagatta, M. (2016). Air pollution prediction via multi-label classification. *Environmental Modelling & Software*, 80, 259–264.
- Costa, P. T., and McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13(6), 653–665.
- Dean, J., Corrado, G. S., Monga, R., Chen, K., Devin, M., Le, Q. V, ... Ng, A. Y. (2012). Large Scale Distributed Deep Networks. *Advances in Neural Information Processing Systems*, 25, 1223–1231.
- Devi, C., Reddy, B., and Kumar, K. (2012). ANN Approach for Weather Prediction using Back Propagation. *International Journal of Engineering Trends and Technology*, 3(1), 19–23.
- Dong, Y., and Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2(1), 222.
- Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*, 12(Jul), 2121–2159.
- Ejimogu, O. H., and Başaran, S. (2017). A systematic mapping study on soft computing techniques to cloud environment. *Procedia Computer Science*, 120, 31–38.
- Espinosa, M. J., and Rodríguez, L. F. G. (2004). Nuestra personalidad: En qué y por qué somos diferentes. Biblioteca Nueva. BIBLIOTECA NUEVA.
- Farnadi, G., Zoghbi, S., Moens, M., & Cock, M. De. (2013). Recognising Personality Traits Using Facebook Status Updates. *Workshop on Computational Personality Recognition (WCPR13) in International AAAI Conference on Weblogs and Social Media (ICWSM13)*, 14–18.
- Gerritsen, L. (2017). Predicting student performance with Neural Networks, (May).
- Gilbert, E., and Karahalios, K. (2009). Predicting tie strength with social media. *In Proceedings of the 27th international conference on Human factors in computing systems - CHI 09 (p.*

- 211). New York, New York, USA: ACM Press.
- Golbeck, J., Robles, C., and Turner, K. (2011). Predicting personality with social media. Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '11, 253.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1), 26–42.
- Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep Learning*. Cambridge: MIT press.
- Hall, J. W. (2017). Examination of Machine Learning Methods for Multi-Label Classification of Intellectual Property Documents. University of Illinois at Urbana-Champaign,.
- Harrington, P. (2012). *Machine learning in action* (Vol. 5). Greenwich, CT: Manning.
- Heaton, J. (2008). *Introduction to neural networks with Java*. Heaton Research, Inc..
- Hedberg, F., Granqvist, I., Nilsson, E., Skjutar, K., and Torstensson, P. (2010). Predicting Team Performance Based on Artificial Neural Networks. In Defining the Future of Project Management. Washington, DC: Project Management Institute. Retrieved from <https://www.pmi.org/learning/library/team-performance-artificial-neural-networks-6494>
- Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., and Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461), 168–174.
- Helwan, A., Idoko, J. B., and Abiyev, R. H. (2017). Machine learning techniques for classification of breast tissue. *Procedia Computer Science*, 120, 402–410.
- Helwan, A., Tantua, D. P., and Adeola, E. (2016). IKRAI: Intelligent Knee Rheumatoid Arthritis Identification. *IJ. Intelligent Systems and Applications Intelligent Systems and Applications*, 1(1), 18–24.
- Humphries, M. (2013). Missing Data & How to Deal: An overview of missing data. Population Research Center. Retrieved from https://liberalarts.utexas.edu/prc/_files/cs/Missing-Data.pdf

- Iatan, I. F. (2017). Predicting Human Personality from Social Media Using a Fuzzy Neural Network. *Issues in the Use of Neural Networks in Information Retrieval*, 661, 81–105.
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... Darrell, T. (2014). Caffe. In *Proceedings of the ACM International Conference on Multimedia - MM '14* (pp. 675–678). New York, New York, USA: ACM Press.
- Kalghatgi, M. P., Ramannavar, M., and Dr. Sidnal, N. S. (2015). A Neural Network Approach to Personality Prediction based on the Big-Five Model. *International Journal of Innovative Research in Advanced Engineering*, 2(8), 56–63.
- Kandias, M., Stavrou, V., Bozovic, N., and Gritzalis, D. (2013). Proactive insider threat detection through social media. In *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society - WPES '13* (pp. 261–266). New York, New York, USA: ACM Press.
- Kanter, J. M., and Veeramachaneni, K. (2015). Deep feature synthesis: Towards automating data science endeavors. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 1–10). IEEE.
- Kee, C. Y., Wong, L.-P., Khader, A. T., and Hassan, F. H. (2017). Multi-label classification of estimated time of arrival with ensemble neural networks in bus transportation network. In *2017 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE)* (pp. 150–154). IEEE.
- Kingma, D. P., and Ba, J. (2015). Adam: A Method for Stochastic Optimization. In *3rd International Conference for Learning Representations*. Retrieved from <http://arxiv.org/abs/1412.6980>
- Kobsa, A. (2007). Generic User Modeling Systems. In *The Adaptive Web* (pp. 136–154). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., and Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, 70(6), 543–556.
- Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable

- from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805.
- Laleh, A., and Shahram, R. (2017). Analyzing Facebook Activities for Personality Recognition. *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 960–964. <https://doi.org/10.1109/ICMLA.2017.00-29>
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Li, H., Su, Y., and Zheng, Z. (2016). Predict Personality with Social Networks.
- Lima, A. C. E. S., and de Castro, L. N. (2014). A multi-label, semi-supervised classification approach applied to personality prediction in social media. *Neural Networks*, 58, 122–130.
- Lison, P. (2012). An introduction to machine learning. Retrieved from <http://folk.uio.no/plison/pdfs/talks/machinelearning.pdf>
- Liu, J., Chang, W.-C., Wu, Y., and Yang, Y. (2017). Deep Learning for Extreme Multi-label Text Classification. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '17 (pp. 115–124)*. New York, New York, USA: ACM Press.
- Liu, S. M., and Chen, J.-H. (2015). A multi-label classification based approach for sentiment classification. *Expert Systems with Applications*, 42(3), 1083–1093.
- Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., and Svetnik, V. (2015). Deep Neural Nets as a Method for Quantitative Structure–Activity Relationships. *Journal of Chemical Information and Modeling*, 55(2), 263–274.
- Maxwell, A., Li, R., Yang, B., Weng, H., Ou, A., Hong, H., ... Zhang, C. (2017). Deep learning architectures for multi-label classification of intelligent health risk prediction. *BMC Bioinformatics*, 18(S14), 523.
- McCrae, R. R., and John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, 60(2), 175–215.
- McGonagle, J., Shaikouski, G., and Hsu, A. (2017). Backpropagation | Brilliant Math & Science Wiki. Retrieved May 3, 2018, from <https://brilliant.org/wiki/backpropagation/>

- Mohammad, S. M., and Kiritchenko, S. (2013). Using Nuances of Emotion to Identify Personality. *In Proceedings of ICWSM (pp. 1–4).*
- Nam, J., Kim, J., Loza Mencía, E., Gurevych, I., & Fürnkranz, J. (2014). Large-Scale Multi-label Text Classification — *Revisiting Neural Networks (pp. 437–452).* Springer, Berlin, Heidelberg.
- Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications, 42(24), 9603–9611.*
- Nielsen, M. A. (2015). Neural networks and deep learning. *Bioinformatics.* Determination Press.
- Nkoana, R. (2011). Artificial Neural Network Modelling of Flood Prediction and Early Warning. UNIVERSITY OF THE FREE STATE BLOEMFONTEIN.
- Ortigosa, A., Carro, R. M., and Quiroga, J. I. (2014). Predicting user personality by mining social interactions in Facebook. *Journal of Computer and System Sciences, 80(1), 57–71.*
- Patel, P. (2017). Why Python is the most popular language used for Machine Learning. Retrieved May 2, 2018, from <https://medium.com/@UdacityINDIA/why-use-python-for-machine-learning-e4b0b4457a77>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12, 2825–2830.* Retrieved from <http://www.jmlr.org/papers/v12/pedregosa11a.html>
- Read, J., Pfahringer, B., Holmes, G., and Frank, E. (2011). Classifier chains for multi-label classification. *Machine Learning, 85(3), 333–359.*
- Santolaya, D. S., and Gavves, E. (2017). Using recurrent neural networks to predict customer behavior from interaction data, (July), 57. Retrieved from <https://esc.fnwi.uva.nl/thesis/centraal/files/f244841390.pdf>
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61, 85–117.*
- SHARMA, S. (2017). Activation Functions: Neural Networks – Towards Data Science. Retrieved May 10, 2018, from <https://towardsdatascience.com/activation-functions-neural->

- Song, H. Y., and Kim, S. Y. (2014). Predicting Human Locations with Big Five Personality and Neural Network. *Journal of Economics*, 2(4), 273–280.
- Statista. (2018). Facebook users worldwide 2017 | Statista. Retrieved April 19, 2018, from <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>
- Sterbak, T. (2017). Guide to multi-class multi-label classification with neural networks in python - Depends on the definition. Retrieved May 13, 2018, from <https://www.depends-on-the-definition.com/guide-to-multi-label-classification-with-neural-networks/>
- Sumner, C., Byers, A., and Shearing, M. (2011). Determining personality traits & privacy concerns from facebook activity. *Black Hat Briefings* 11, 1–29.
- Tabatabaei, S. M., Dick, S., and Xu, W. (2017). Toward Non-Intrusive Load Monitoring via Multi-Label Classification. *IEEE Transactions on Smart Grid*, 8(1), 26–40.
- Tandera, T., Hendro, Suhartono, D., Wongso, R., and Prasetio, Y. L. (2017). Personality Prediction System from Facebook Users. *Procedia Computer Science*, 116, 604–611.
- Tsoumakas, G., and Ioannis, K. (2006). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining* 3.3, 3(3).
- Wald, R., Khoshgoftaar, T. M., Napolitano, A., and Sumner, C. (2012). Using Twitter Content to Predict Psychopathy. In *2012 11th International Conference on Machine Learning and Applications* (pp. 394–401). IEEE.
- Wilson, R. E., Gosling, S. D., and Graham, L. T. (2012). A Review of Facebook Research in the Social Sciences. *Perspectives on Psychological Science*, 7(3), 203–220.
- Wong, T.-T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, 48(9), 2839–2846.
- Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K. C., ... Frey, B. J. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science (New York, N.Y.)*, 347(6218), 1254806.

- Zhang, J. (2016). Deep learning for multi-label scene classification by Declaration of Authorship, (August).
- Zhang, M.-L. Z. M.-L., and Zhou, Z.-H. Z. Z.-H. (2006). Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization. *IEEE Transactions on Knowledge and Data Engineering*, 18(10), 1–14.
- Zhang, M.-L., & Zhou, Z.-H. (2007). ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7), 2038–2048.
- Zhang, W., Yan, J., Wang, X., and Zha, H. (2017). Deep Extreme Multi-label Learning. Retrieved from <http://arxiv.org/abs/1704.03718>

APPENDICES

SOURCE CODE

- Merging database

The screenshot shows the SQL Server Enterprise Manager interface on the left, displaying the database structure for 'CISLAB218-1'. The main window shows a SQL query in 'SQLQuery8.sql' that performs a merge operation. The query uses COALESCE to handle missing values and FULL JOIN to merge data from two tables. The results pane shows a table with 12 rows and 10 columns: 'userid', 'ope', 'neu', 'agr', 'con', 'ext', 'age', 'relationship_status', 'gender', and 'network_s'.

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT COALESCE(b.[userid"], d.[userid"], f.[userid"]) [userid"]
,MAX(b.[ope"]) ope, MAX(b.[neu"]) neu, MAX(b.[agr"]) agr, MAX(b.[con"]) con,MAX(b.[ext"])
,MAX(d.[age"]) age, MAX(d.[relationship_status"]) relationship_status, MAX(d.[gender"]) ger
,MAX(f.[n_like"]) n_like, MAX(f.[n_status"]) n_status, MAX(f.[n_event"]) n_event, MAX(f.[r
FROM [new merged].[dbo].[big5] b
FULL JOIN demog d ON b.[userid"] = d.[userid"]
FULL JOIN freq f ON b.[userid"] = f.[userid"]
WHERE ["n_like"] IS NOT NULL
AND ["gender"] IS NOT NULL
GROUP BY COALESCE(b.[userid"], d.[userid"], f.[userid"])
ORDER BY COALESCE(b.[userid"], d.[userid"], f.[userid"])

```

	"userid"	ope	neu	agr	con	ext	age	relationship_status	gender	network_s
1	"000005636dcd002ae6689e8097df7"	"3.50"	"2.75"	"4.00"	"4.50"	"4.67"	"37"	""	"1"	"157"
2	"0000130571654e3afaa62f4e9d2e4f63"	"3.00"	"4.25"	"2.75"	"2.75"	"4.00"	""	"2"	"0"	"193"
3	"00001627067cf12b1923f02bb1a3b731"	"4.15"	"3.10"	"3.95"	"4.15"	"4.65"	"20"	"2"	"1"	"84"
4	"0000164ed0ba466826f8aa1abd1805c"	"3.70"	"2.80"	"3.70"	"3.30"	"3.50"	"21"	"1"	"0"	"56"
5	"00001cd4d03dfe8af01e9bb08af119b8"	"3.80"	"2.05"	"3.25"	"3.95"	"3.00"	"28"	"2"	"1"	"37"
6	"00002503c9e302d6972ef427e0abb18"	"3.75"	"3.25"	"3.50"	"2.75"	"4.00"	""	""	"0"	"124"
7	"00002a342c30e272f96a5f46d6b0f42a"	"4.25"	"1.25"	"2.50"	"4.25"	"4.25"	"21"	"2"	"1"	"547"
8	"00002da5f6ccb871cf62c3364fbaeeb8"	"4.45"	"3.45"	"3.50"	"2.40"	"2.85"	"20"	"1"	"0"	"289"
9	"0000370358b5385427f3979fa7a4d35c"	"3.00"	"2.38"	"2.75"	"3.75"	"2.88"	"30"	"1"	"0"	"238"
10	"0000481d9f676d183ed85a90f5f27356"	"4.50"	"1.00"	"5.00"	"5.00"	"3.25"	"20"	"2"	"0"	"489"
11	"00004c5fbc24c3f800b13974ee7ab18e"	"3.50"	"1.75"	"3.50"	"4.00"	"3.50"	""	"1"	"1"	"464"
12	"000052th6030cf407hc8f982632hee17"	"4.50"	"2.00"	"3.75"	"2.88"	"4.63"	""	""	"0"	"1176"

Query executed successfully. CISLAB218-1 (13.0 RTM) CISLAB218-1\cislab (57) new merged 00:00:42 1337313 rows

- Dealing with missing values

The screenshot shows a SQL query in 'SQLQuery1.sql' that selects all columns from a table named 'Merged Data1' and then deletes rows where the 'n_like' column is NULL or empty.

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT *, ["userid"]
, ["ope"]
, ["neu"]
, ["agr"]
, ["con"]
, ["ext"]
, ["gender"]
, ["age"]
, ["network_size"]
, ["relationship_status"]
, ["n_like"]
, ["n_status"]
, ["n_event"]
, ["n_group"]
, ["n_tags"]
FROM [merged].[dbo].[Merged Data1]
delete from [Merged Data1] where ["n_like"] IS NULL OR ["n_like"]="

```

- **Code for importing libraries using keras framework**

```
import tensorflow as tf
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import neurolab as nl
import pandas as pd
import xlswriter
from sklearn.preprocessing import Normalizer
from sklearn.model_selection import StratifiedKFold
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import Dense, Dropout, Activation
from sklearn.preprocessing import MultiLabelBinarizer
from keras.optimizers import SGD
from keras.wrappers.scikit_learn import KerasClassifier
from keras.utils import np_utils
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import KFold
from sklearn.preprocessing import LabelEncoder
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
from sklearn.preprocessing import Binarizer
from sklearn.preprocessing import MinMaxScaler
from scipy import optimize
%matplotlib inline
```

- **Importing Dataset into python using pandas**

```
dataframe = pd.read_csv("sampNNdata.csv", header=None)
dataset = dataframe.values
X = dataset[:,0:9].astype(float)
y = dataset[:,9:]
```

- **One Hot Encoder to create dummy variables**

```
onehotencoder = OneHotEncoder(categorical_features = [1])
X = onehotencoder.fit_transform(X).toarray()
X = X[:, 1:]
```

- **Normalizing the Data**

```
normy = y/np.amax(y, axis=0)

rescaler = MinMaxScaler(feature_range=(0,1))
rescaledX = rescaler.fit_transform(X)
mlb = Binarizer(threshold=0.5).fit(normy)
mlb_y=mlb.transform(normy)
```

- **Creating the neural network model and K-fold validation**

```
from sklearn import metrics
from sklearn.metrics import accuracy_score
from keras import callbacks
from keras import optimizers
import time

# Fitting our model
earlyStopping=callbacks.EarlyStopping(monitor='val_loss',
                                     min_delta=0,
                                     patience=10,
                                     verbose=0, mode='min')
```

```
optimizers.Adam(lr=0.0001, beta_1=0.9, beta_2=0.999, epsilon=None, decay=0.0,
amsgrad=False)
```

```
# define 10-fold cross validation test harness
```

```
seed = 7
```

```
np.random.seed(seed)
```

```
kfold = KFold(n_splits=5, shuffle=False, random_state= seed)
```

```
cvscores = []
```

```
start = time.time()
```

```
for train, test in kfold.split(rescaledX, mlb_y):
```

```
    from keras import regularizers
```

```
    classifier = Sequential()
```

```
    # Adding the input layer and the first hidden layer
```

```
    classifier.add(Dense( input_dim = 18,
                          init = 'uniform', output_dim = 30, activation='relu'))
```

```
    # Adding the second hidden layer
```

```
    classifier.add(Dropout(0.3))
```

```
    classifier.add(Dense(output_dim = 5, activation = 'sigmoid'))
```

```
    classifier.compile(optimizer = 'Adam', loss = 'binary_crossentropy', metrics =
['accuracy'])
```

```
    history = classifier.fit(rescaledX[train], mlb_y[train], validation_data=(rescaledX[test],
mlb_y[test]), epochs=1000, batch_size=30, callbacks=[earlyStopping])
```

```
# evaluate the model
```

```
    scores = classifier.evaluate(rescaledX[test], mlb_y[test])
```

```
    print("%s: %.2f%%" % (classifier.metrics_names[1], scores[1]*100))
```

```
    cvscores.append(scores[1] * 100)
```

```
print("%.2f%% (+/- %.2f%%)" % (np.mean(cvscores), np.std(cvscores)))
```

```
# Fitting our model
```

```

#history = classifier.fit(X_train, Y_train, validation_data=(X_test,Y_test), batch_size =
50, nb_epoch = 200)

# show the accuracy on the testing set
print("[INFO] evaluating on testing set...")
(loss, accuracy) = classifier.evaluate(rescaledX[test], mlb_y[test],
    batch_size=10)
print("[INFO] loss={:.4f}, accuracy: {:.4f}%".format(loss,
    accuracy * 100))
end = time.time()
print ("Model took %0.2f seconds to train"%(end - start))

```

- **Plotting the Results**

```

# list all data in history
print(history.history.keys())
# summarize history for accuracy
plt.plot(history.history['acc'])
plt.plot(history.history['val_acc'])
plt.title('model accuracy')
plt.ylabel('accuracy')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.savefig('accuracy.png')
plt.show()
# summarize history for loss
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')


















```

```
plt.legend(['train', 'test'], loc='upper left')
plt.savefig('loss.png')
plt.show()
```

- **Predicting**

```
from sklearn.metrics import accuracy_score
# predict
y_pred = classifier.predict(rescaledX[test])
predicted = Binarizer(threshold=0.5).fit(y_pred)
predicted_y=predicted.transform(y_pred)

print (y_pred)
print (predicted_y)
accuracy2 = np.mean(predicted_y == mlb_y[test])
print("Prediction Accuracy: %.2f%%" % (accuracy*100))
```

<input type="checkbox"/>	AUTHOR	TITLE	SIMILARITY	GRADE	RESPONSE	FILE	PAPER ID	DATE
<input type="checkbox"/>	Obima H. Ejimogu	Abstract	0% 	--	--		971706467	03-Jun-2018
<input type="checkbox"/>	Obima H. Ejimogu	Conclusion	0% 	--	--		971706448	03-Jun-2018
<input type="checkbox"/>	Obima H. Ejimogu	Introduction	0% 	--	--		971706465	03-Jun-2018
<input type="checkbox"/>	Obima H. Ejimogu	Methodology	3% 	--	--		971706455	03-Jun-2018
<input type="checkbox"/>	Obima H. Ejimogu	Literature Review	5% 	--	--		971706461	03-Jun-2018
<input type="checkbox"/>	Obima H. Ejimogu	Results	6% 	--	--		971706452	03-Jun-2018
<input type="checkbox"/>	Obima H. Ejimogu	Conceptual Framework	9% 	--	--		971706458	03-Jun-2018
<input type="checkbox"/>	Obima H. Ejimogu	 Thesis	16% 	--	--		971706451	03-Jun-2018