PREDICTING THE DEPRESSION LEVEL OF EXCESSIVE USE OF MOBILE PHONE: USING MACHINE LEARNING ALGORITHM

BY

Imrus Salehin
ID: 171-15-8978

AND

Iftakhar Mohammad Talha
ID: 171-15-9019

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Ms. Nazmun Nessa Moon
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Farah Sharmin
Sr. Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

31 MAY, 2021

APPROVAL

This Project titled **"Predicting the Depression Level of Excessive Use of Mobile Phone: Using Machine Learning Algorithm"**, submitted by **Imrus Salehin, ID No: 171-15-8978** and **Iftakhar Mohammad Talha, ID No: 171-15-9019** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **31 May, 2021.**

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Gazi Zahirul Islam — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Raja Tariqul Hasan Tusher — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Dewan Md. Farid — **External Examiner**
Associate Professor
Department of Computer Science and Engineering
United International University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Ms. Nazmun Nessa Moon , Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Ms. Nazmun Nessa Moon
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Farah Sharmin
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Imrus Salehin
ID: 171-15-8978
Department of CSE
Daffodil International University

Iftakhar Mohammad Talha
ID: 171-15-9019
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty Allah for divine blessing makes us possible to complete the final year Thesis Project successfully.

We really grateful and wish our profound our indebtedness to **Ms. Nazmun Nessa Moon, Assistant Professor,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning"* to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this Thesis project.

We would like to express our heartiest gratitude to **Mohd. Saifuzzaman**, **Farah Sharmin**, and Head **Professor Dr. Touhid Bhuiyan**, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

In this research titled “Predicting the Depression Level of Excessive Use of Mobile Phone: Using Machine Learning Algorithm” which is applied advanced machine learning and regression analysis to find out the depression level. We have done the whole work in the research area of medical science and information technology and also built up a collaboration. In this study, we are focusing on the strength of the algorithm and also calculate the accuracy with python programming. The result expresses that smart mobile device changing the human brain day by day if spend more time around 8 to 12 hours a day. At last, we observed that a man or woman slowly going through a depression for the impact of the excessive mobile operates. In our study, we have used multiple classification algorithms to find out depression level such as Probability, Decision Tree, Random Forest, Linear Regression and SVM (Support Vector Machine). For the accuracy of our work, we have used four types of algorithms to find the optimal ratio and percentage.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Depression is a temporal change in the human brain that can slowly make people clinically imbalanced and if this imbalance is caused by an addiction to a mobile device then it can be fatal. Depression reduces feelings of sadness or interest in their own activities [1]. Because this generation is slowly getting no idea about the right technology, they are unaware of the magnitude and breadth of its use. This perception can be a major threat that can later be described as a mental disorder. We are using a framework for analyzing data in raw data mining research [2]. Ignorance of the use of the Mobile device and for the cause of extreme technologies use to fall in depression, we have tried to establish in this paper the relation between these issues. We have done this whole work in the areas of data mining and machine learning.

1.2 Motivation

In this thesis, we are trying to look at the social, emotional, health effects and some bad effects of mobile phone use. These are given below:

- Bad effects of extreme using the mobile phone
- Mental addiction of using mobile phone
- Long-time depression disorder
- Build up a collaboration between ICT and Medical Sector
- Smart phone Side effect in social life and personal life

1.3 Hypothesis of the Study

We are focusing on a more interesting and meaningful question, such as psychological questions; Timing of using the device; Felling addiction, etc. For the main structure, we are using a "Tree based structure", that's why we can easily represent it in a clear way [3]. In this paper, we propose the best framework for prediction of Depression quality and level. Another side, we mainly focus on the accuracy of the depression level by using international scale value. We are measuring the value using data mining tables, graphs, formulas, and algorithms also. After completing all the process, we test our probability of accuracy result and give a percentage decision about the time factor and depression level.

1.4 Expected Output

The main contribution of the research which is summarize as the following:

- As we have studied, excessive mobile use is a major cause of imbalance in the human brain. We identify it using a supervised machine learning algorithm and data mining
- In this research, we are working with the Time factor and the Human brain Nero molecular system. This system is implemented with Machine learning and Decision tree.
- The proposed hybrid methodology, human depression level prediction is the most significant for this study. Decision tree classifier, python coding, Random Forest and SVM (Support Vector Machine) algorithm, Heat map are the major function in this study which are create an accuracy.
- In this research, we are describing the combination of the medical sector and machine learning. We are focusing on the harmful and mental disorder prediction which is very vital to our upcoming world and society.

CHAPTER 2
BACKGROUND

2.1 Introduction

Nowadays the number of smartphone users is increasing day by day. This is why we focus on this issue which is very sensitive to human beings. In the past, many scholars have done research on the effects of mobile but we are trying to find out about the level of the human brain and depression and its effects. This is a sensitive issue. We don't take it seriously, but because of depression, we can face many problems physical and mental. It can harm us in many ways. Depression can lead to suicidal thoughts or feelings of inadequacy, personal unrest or conflict with family, physical abuse, sleep problems, chronic pain, anxiety, etc. Adolescents are more at risk of depression, so it is more dangerous.

2.2 Literature Review

In this study, they proposed a model based on real-time data from Twitter. Sleepiness, erotic thoughts, restlessness is related to anxiety disorders. They created a 5 tuple vector and those vectors are word, timing, frequency, sentiment, and contrast. The timing and frequency of the tweets are analyzed. They use Multinational Navigation Base, Gradient Boosting, and Random Forest Classification to train their models and they achieve 85.09% accuracy [4]. In this study, they try to find out how they can use internet-based treatment for depression. In their study, they focus on this internet-based knowledge that they can use and apply. How can they use it in mental health care, general treatment fields, etc ? [5]. By 2030 depression will become a serious problem and cause of disability. According to the World Health Organization. This document says that, what is the role of inflammatory cytokines in depression [6]. Correlation analysis is a widely used technique to find out the relationship between data. In this study, they used correlation and machine learning to see the patient's mental health conditions. They use WEAKA tools for correlation. They have depressive disorder symptoms dataset and situations based on weather datasets [7]. This paper presents a method to analyze emotions. They proposed a two-step approach. Firstly identify the emotional words and then using an algorithm they extract the values of those emotional words to provide the result [8]. This research shows that the higher the usage of social media, the higher the risk of depression, with teenage girls being subjected to the highest risk. An early depression detector is proposed to track and control this risk factor of social media usages [9]. Another paper introduces how internet addiction causes depression and affects in adolescents and emerging adults. They have 780 people's datasets. They have two groups. 390 are adolescents and their ages 12-17 years and 192 females. Another group consists of 328 adults and their ages 18–30 years and 197 females. The results show that the Internet involved in depression is addictive [10]. Artificial Neural Network (ANN) is used to predict the strength of self-compacting concrete. They use 99 data samples for their work. ANN‟s is considered as nonlinear statistical data modeling tools. In this study they discuss about how to perform prediction using the ANN (Artificial Neural Network) [11]. In this paper, Miami health program named as Resources for Enhancing Alzheimer's Caregiver Health observe the ability of family therapy and technology-based intervention in reducing depressive symptoms. How family therapy can affect how it can reduce depression they examined it. They examined their patients at 6 months and 18 months follow-up [12]. In this paper, a long survey ranging from 2001 to 2002 was conducted. They use the internet. All the participants use the internet for communicating with their friends and their families. Six months later they showed a lower depression [13]. In this study, they try to find out how technology is related to depression, sleep quantity, anxiety, time to awake. 236 college students give their data for this survey. Regression analysis tells that a higher level of technology use causes depression, causes poorer sleep quality etc. [14].

2.3 Research Summary

We proposed hybrid methodology, algorithm strength identification and human depression level prediction is the most significant for this study. Decision tree classifier, python coding, Random Forest, Support Vector Machine algorithm, and Heat map are the major function in this study which are create an accuracy. We are describing the combination of medical sector and machine learning. We are focusing the harmful and mental disorder prediction which is very vital to our upcoming world and society. Technological device or mobile extreme use and its Impact of the Human Brain.

2.4 Challenges

Our research will have some challenges:

- Collecting a lot amount of data.
- After collecting manage the data.
- Data procedure is our main problem in this research.
- Experimental result and outcome result also a problem.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

This chapter presents the methodology that was used in the study of the impact of mobile phones on predicting the depression disorder. It begins Research Subject and Instrumentation, Data Collection Procedure, Statistical Analysis, and Implementation Requirements.

3.2 Research Subject and Instrumentation

Research instruments were used and played a key role in guiding the researcher into choosing a combination of qualitative and quantitative research methods. The questionnaire is a critical stage in the survey research process, the questionnaire must be relevant and accurate in trying to capture the essence of the research objective. To achieve these ends, a research will be required to make several decisions:

- Which data should be collected?
- How should each question be phrased?
- In what sequence should the questions be arranged?
- How to ensure that the collected data are okay?
- Does the questionnaire need to be revised?

3.3 Data Collection Procedure

Data is the core of each research and analysis. Data collection always allows us to gather the information needed for any kind of analysis. Depending on the type of research data, document review data, Observation data, Questioning data, measurement data, or a combination of different methods data. To collect our data we use online data, survey data from our friends and family, medical data, providing google form for online participants and university student's data.

- **Questionnaire**

Data collection based on questionnaires is a way that consists of a series of questions and participants provide feedback based on these questions.

Steps required to design Questionnaire

1 Defined the objectives of the study

2 Define the target audience, methods and ways to reach them

3 Question Design

4 Question Testing

5 Questionnaire Administration

6 Results Interpretation

3.3.1 Data Preprocessing & Organizing

Data pre-processing refers to the pre-phase of processing the dataset. Generally, raw data sets are not able to perform operations and generate the expected outcomes. As a result, data Pre-Processing is required and it is considered to be one of the most important parts of research. The questionnaire contained four independent pieces of information to be analyzed through machine learning. We have some string data in our dataset. Using the string data we can't get the proper output that we have needed. In our dataset, we have 2 columns that contain string values. So, we converted those string values into numeric values in 2 steps shown in Figure 3.1 and 3.2. At first, we converted the "Time spend with devices column". We named the new converted column as "time_spend_numeric" and specified the values as 0, 1, 2, 3, 4.

Where,

0 = "6 hour"

1 = "less than 12 hour"

2 = “less than 1 hour”

3 = “less than 8 hour”

4 = “more than 12 hour”

```
In [16]: #converting string data to numerical data
         def trans_time_spend(x):
             if x == '6 hour':
                 return 0
             if x == 'less than 12 hour':
                 return 1
             if x == 'less than 1 hour':
                 return 2
             if x == 'less than 8 hour':
                 return 3
             if x== 'more than 12 hour':
                 return 4
```

Figure 3.1: Converting String Data to Numeric

	Sex	Age	device_use	feel_loss	having_meal	lonely	xclass	time_spend_numeric
0	1	17	1	0	0	0	FD	1
1	1	26	0	1	0	0	MD	3
2	2	18	1	0	0	1	LD	3
3	1	19	1	0	0	1	MD	3
4	1	20	1	0	0	1	LD	3
5	2	21	1	0	1	1	LD	3
6	2	22	1	1	1	0	LD	3
7	1	23	1	0	1	0	LD	3
8	1	19	1	0	1	0	LD	3
9	1	20	0	0	0	0	FD	1

Figure 3.2: Data Set after Converting

Now we have also one column left to convert from string to numeric which is shown in Figure 3.3 and 3.4. This is our predicted class which is named xclass. We classify this xclass data into 3 categories which are LD(Low Depression), MD(Mid-level Depression), FD(Full Depression). Also here we defined the values as 0, 1, 2.

Where,

0 = “LD”

1 = “MD”

2 = “FD”

```
In [19]:
def trans_class(x):
    if x == 'LD':
        return 0
    if x == 'MD':
        return 1
    if x == 'FD':
        return 2
```

Figure 3.3: Converting xclass String Data to Numeric Data

	Sex	Age	device_use	feel_loss	having_meal	lonely	time_spend_numeric	class
0	1	17	1	0	0	0	1	2
1	1	26	0	1	0	0	3	1
2	2	18	1	0	0	1	3	0
3	1	19	1	0	0	1	3	1
4	1	20	1	0	0	1	3	0
5	2	21	1	0	1	1	3	0
6	2	22	1	1	1	0	3	0
7	1	23	1	0	1	0	3	0
8	1	19	1	0	1	0	3	0
9	1	20	0	0	0	0	1	2

Figure 3.4: Numeric Dataset

3.4 Model Analysis

We use this tree-based system for prediction and calculate the accuracy for our model. The first node is called the root node and this node gets divided. Dividing a node into sub-nodes is the process of splitting. The leaves are the nodes when the nodes can no longer divide. Figure 3.5, shows the basic idea that how the algorithm works smoothly. For the first step, we need the dataset. It will select the best feature for segmentation using the Attribute Selection Measures. Then it will make that feature, the decision node. The next step is to break down the data set and divide it into 2 subsets, one is the training dataset and the other one is test dataset. Now the tree-building process will start. This process will continue until one of the conditions will match. The next step is to test the data with the pre-training dataset with the evaluated model. At last, the model will give the accuracy of the entire work.

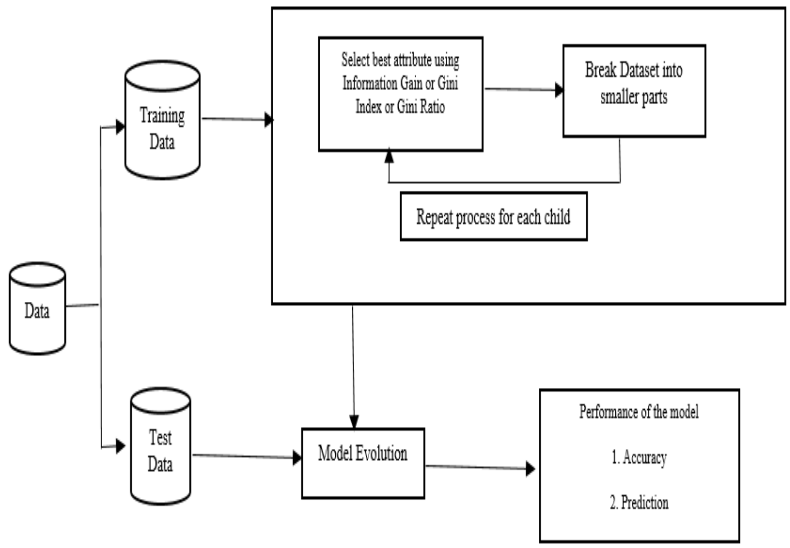

Figure 3.5: Working Process of the Model

We have to initialize all the library first. We divided our the dataset at 80% of the training data and 20% of the test data. Test data will be selected randomly for better prediction. Decision Tree process will continue using true and false measures. When it finds the leaf node the splitting procedure will stop. If it found a false value, the splitting procedure will also stop then and also give a level for depression measure. We divide those levels as non-clinical or not depressed level, Low-level depression, Mid-level depression, and Major depression or fully depressed shown in Figure 3.6. This arrangement will help us a lot, such as how desperate someone is and how dire their situation is.

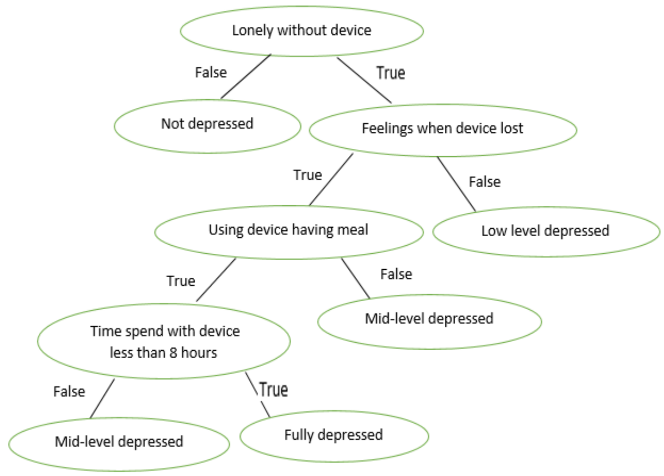

Figure 3.6. Tree Implementation Using the Questionnaires

3.5 Data Category and Brain Analysis Data Criteria section

To determine depression, the possible symptoms of depression and its level maintain some large rules and categories shown in Table 3.1 [15, 16]. Technology is responsible for the frustration or depression, which we have divided into several sections.

Table 3.1: Determine the Depression Level Category

Category	Criteria
Major Depression	If accuracy level is stay between 60-75 % medical sector say it will be going to in depression. Time factor must be depend for change the brain chemicals function
Mid-Level Depression	If accuracy level is stay between 40-55 % medical sector say it will be going to in Mid-Level depression. Time factor must be depend for change the brain chemicals function
Low Depression	If accuracy level is stay between 35% + medical sector say it will be going to in Low depression. Time factor must be depend for change the brain chemicals function
Non Clinical	Below 30% accuracy level, we cannot say a person affected with depression. It is normal chemical reaction of brain.

3.6 Statistical Analysis

For the statistical analysis we have used Statistical Package for the Social Sciences SPSS and Microsoft excel for data entry and analysis. Pearson's correlation technique tool also used to establish the relationships among the variables.

3.6.1 Male & Female Participants Result

In this research we collect the different types of aged people for calculation and accuracy. We collect data from many people and a special organization. This data was chosen very carefully and randomly. We ensure the privacy and confidential information very carefully also in our survey; we get 56% male participation and 44% of female participation, which is shown, in Table 3.2. Figure 3.7 represents the counter plot diagram of male and female.

Table 3.2: Total Data in Age

Age Group (Year)	Person Type	Male	Female
8-12	Children	15	12
13-18	Teenager	231	169
19-30	Younger	587	485
31-above	Old	86	63
Total		= 919	= 729

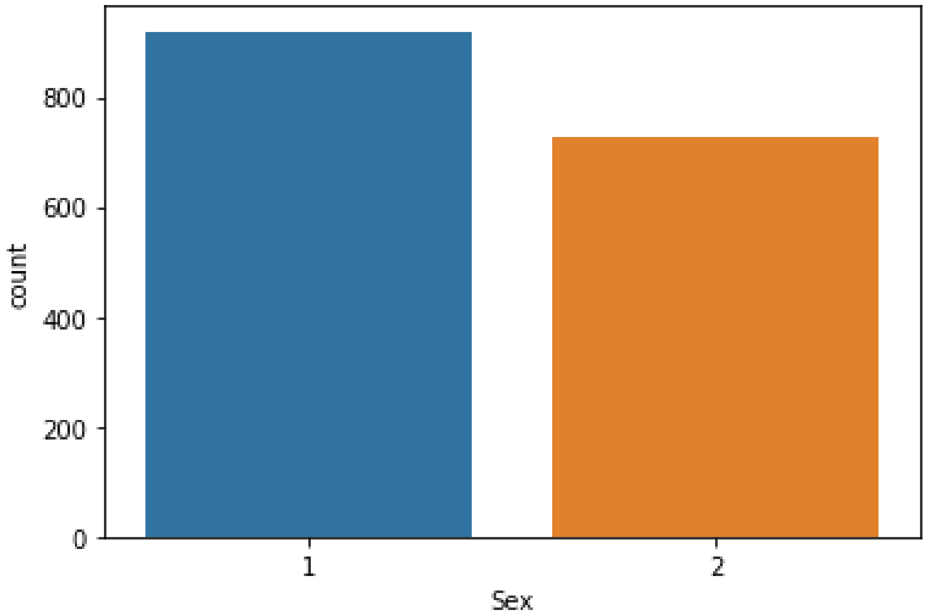

Figure 3.7: Male Female Counter Plot

3.6.2 Device Used by the Participants

Table 3.3: Device Respondent

Device	Frequency
Mobile	1127
Laptop	426
Desktop	95
Total	= 1648

From Table 3.3, we can observe that from our participants 1127 respondents use mobile phones, 426 respondents use laptop, and 95 respondents use desktops which is shown in Figure 3.8.

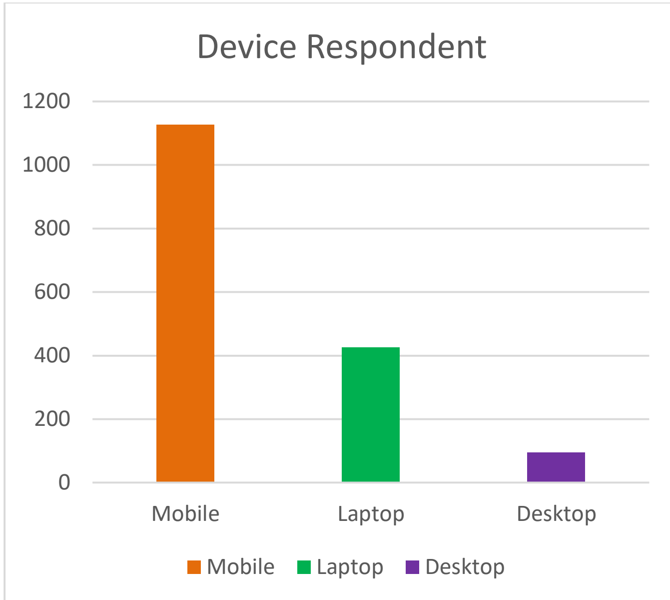

Figure 3.8: Device Respondent Plot

3.7 Implementation Requirements

- **Algorithm**

To complete our work, we have used decision tree, Random forest and Support vector Machine algorithm for finding out the accuracy of our work. We find decision tree is the most effective algorithm for our research. We have used the Attribute Selection Measure or ASM technique to select the best attributes of the root node. To perform the ASM method we used the Gini Index technique. Splitting nodes chooses the best value for the best partition and a random value for the random partition. However, it takes the best value for the division. This can be achieved using the Gini Index splitting process.

$$Gini = 1 - \sum_{i=1}^{c} (pi)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \; 1$$

The Gini Index measures the probability or degree of incorrectness of the variables. When a variable is chosen randomly, it can be wrong or misclassified. The Gini index operates between 0 and 1 where 0 indicates that all elements are in the same category, and 1 indicates that all elements are chosen from the different categories randomly.

- **Hardware and Software**
  - **Python 3.7:-** Python 3.7 is a Python version. It is a high-level programming language. Most of the researchers use it to do their research.
  - **Google Colab:-** Google Collab is a free to use open-source distributor of Python programming language.
  - **SPSS:-** We have used SPSS (Statistical Package for the Social Sciences) for statistical analysis and data entry.
  - **Operating System:-** Windows (7, 8.1 or 10).
  - **Browser:-** Google Chrome.
  - **RAM & ROM:-** Hard Disk (Minimum 4 GB), ROM(Minimum 4 GB)

CHAPTER 4
EXPERIMENTAL RESULT AND DISCUSSIONS

4.1 Introduction

In this section we discussed about the results and data analysis. The result and data analysis is based on the research objectives. In order to show the output results and data analysis report, tables and graphs are used.

4.2 Experimental Results

Depression and brain both are interconnected. Excessive time spent with the device increases the amount of depression. The amount of time spent with the device will increase the amount of frustration. In this section we determine the relationship of depression to our questionnaire and discuss the results obtained through algorithms.

4.2.1 Favorite Device Loss Feeling Respondent

What will happen to them when someone loses their favorite device like mobile phone, laptop, smart watch etc. This feeling of loss can cause depression, from Figure 4.1, we can very easily illustrate it. The level of depression is quite high when someone loses their favorite device.

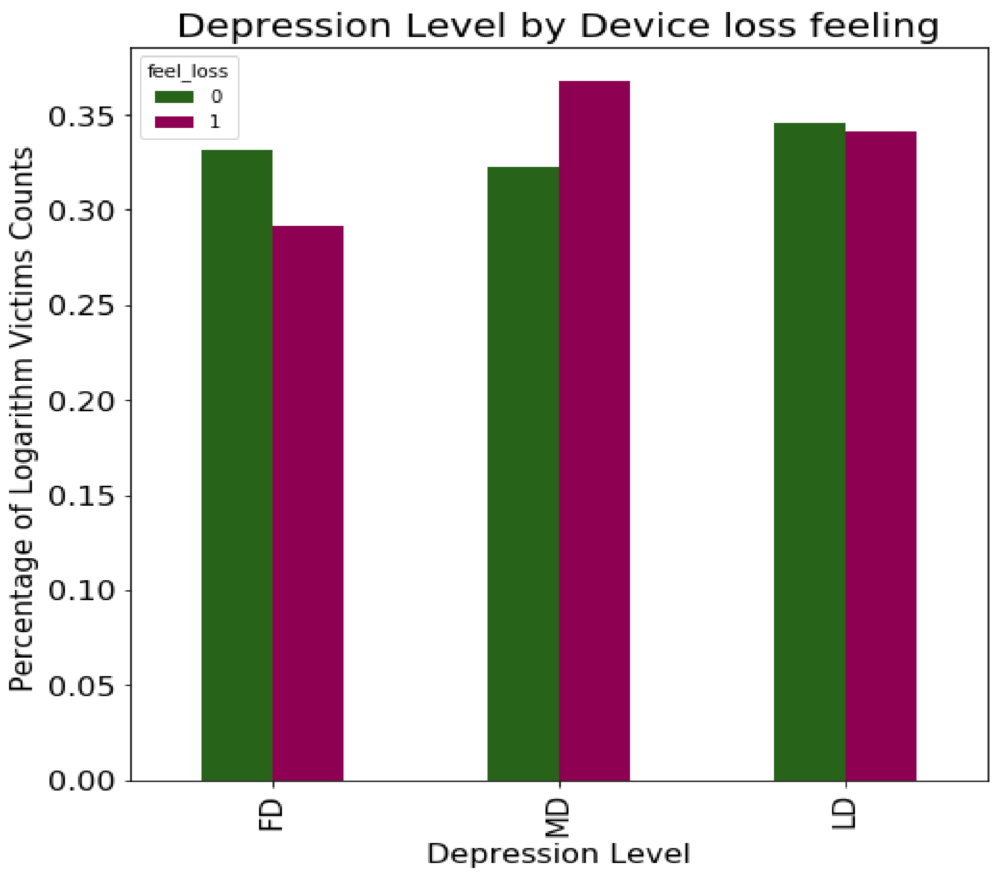

Figure 4.1: Favorite Device Loss Feeling

4.2.2 Depression Level by Time Spending with Device

Those who spend a lot of time with devices have a very high level of frustration. Through our studies we have found that those who spend at least 8 to 12 hours using the device will have a much higher level of depression than those who spend less time with the device shown in Figure 4.2.

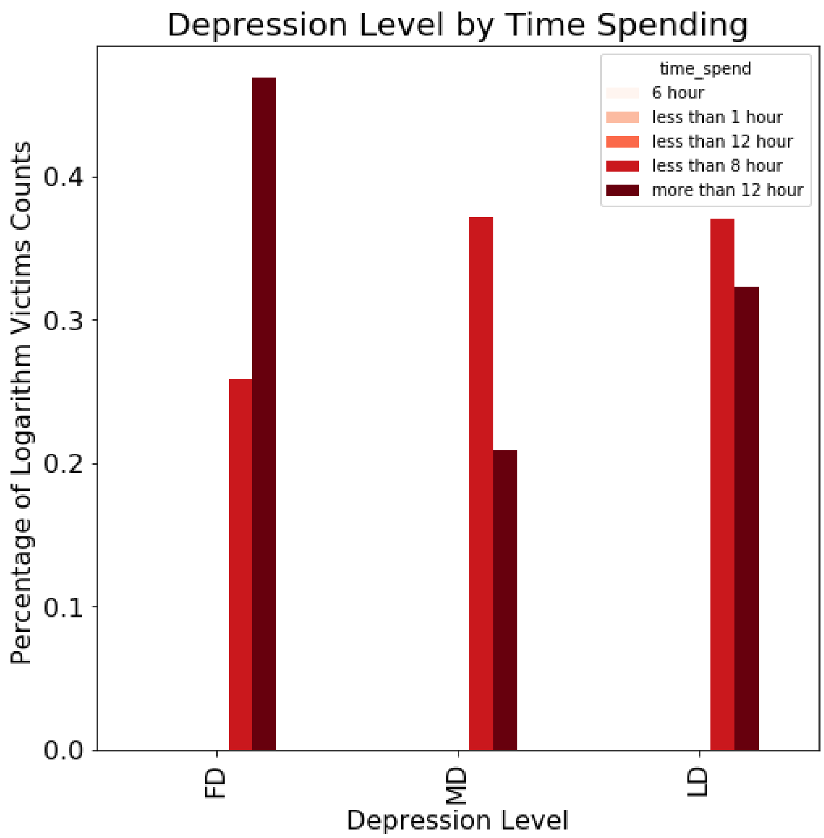

Figure 4.2: Depression Level by Time Spending with Device

4.2.3 Depression Level by Having Meal with Devices

Depression is a major cause of failure. Those who are depressed can do nothing fruitful in life. People who are attached to the device while eating, or who are busy with the device while talking to parents and friends, may also have a higher level of frustration shown in Figure 4.3. Because they are more addicted with device they are less connected with the real world.

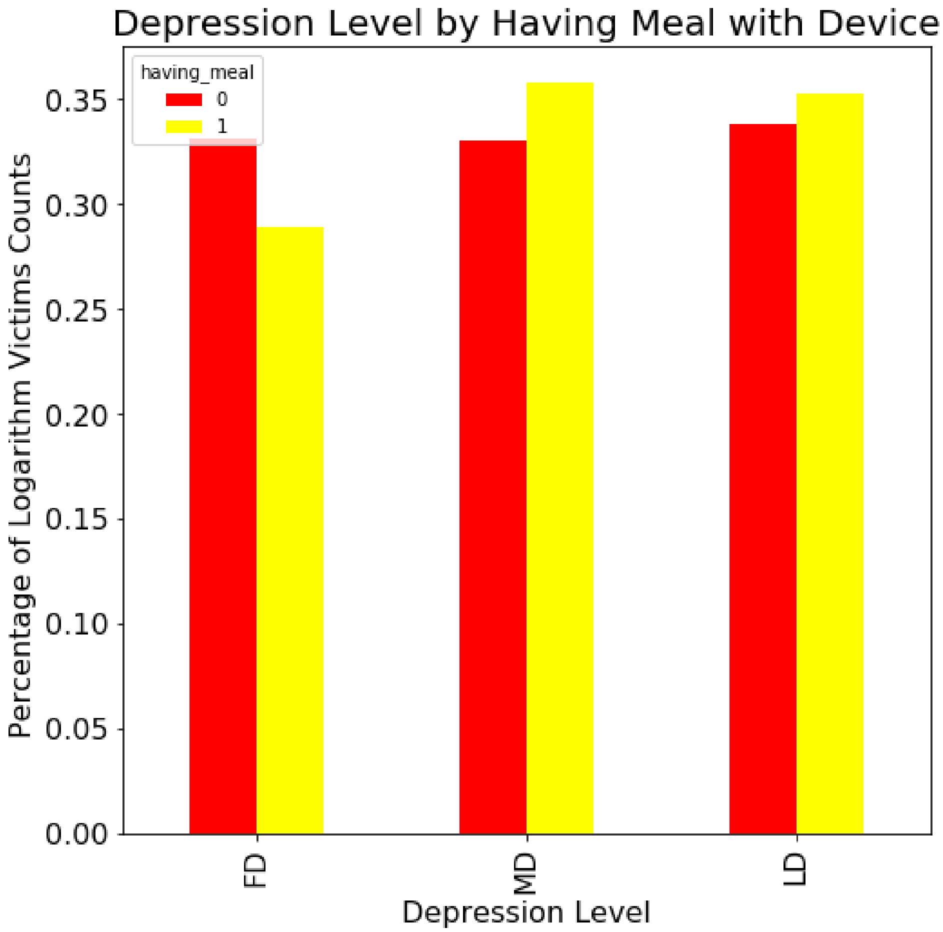

Figure 4.3: Depression Level by Having Meal with Device

4.2.4. Depression level by lonely feeling without device use

People, who think that they are very lonely without device, those who think only phones or laptops or other devices can make them happy, make them more confident, definitely their level of depression is at high stage. Figure 4.4, Depression Level by Lonely Feeling without Device

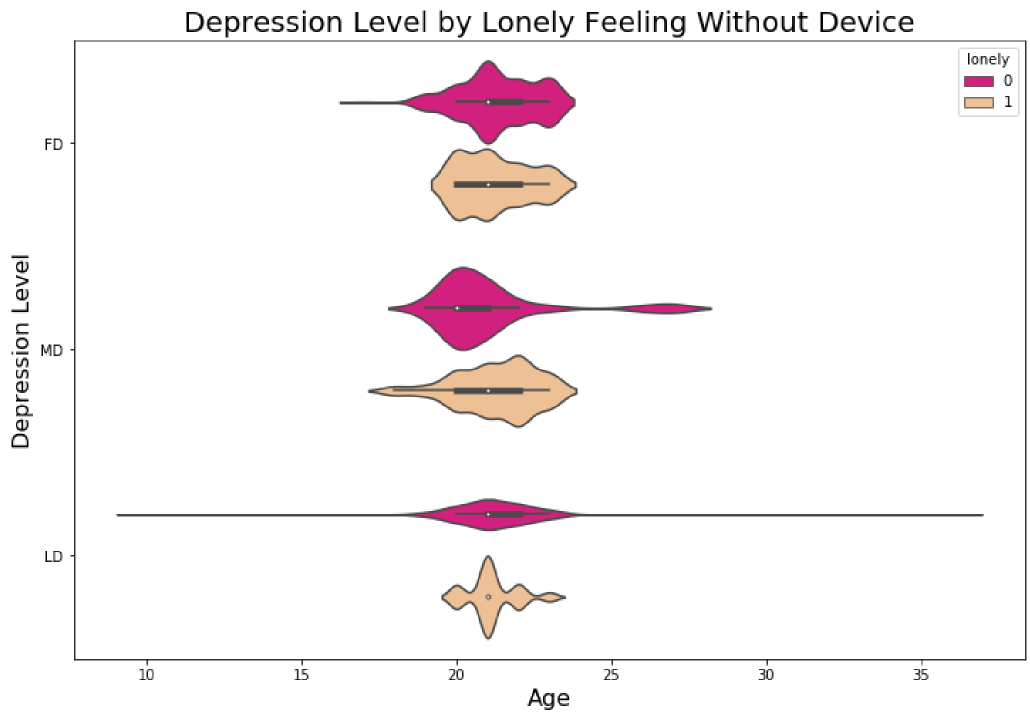

Figure 4.4. Depression Level by Lonely Feeling without Device

4.3 Implementation of Interactions

In this part we will discuss how we evaluate our depression prediction model. For prediction we use machine learning algorithm with the Google Colab IDE. Whole process is given below:

- **Importing Required Library functions**

For the first step, we imported our required library functions to complete this study which is shown in Figure 4.5.

```
#required library import

%matplotlib inline

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

import statsmodels.api as sm

import pandas.util.testing as tm
```

Figure 4.5: Importing All Required Library Functions

- **Dataset Load and visualization**

Then we load our dataset which we have collected from the respondent shown in Figure 4.6.

```
#dataset import
df = pd.read_csv(r"E:\Cse Daffodil\Paper\Technology on depression\Defns data set\data_test_22.csv")
df.head(10)
```

	Sex	Age	device_use	feel_loss	time_spend	having_meal	lonely	COND1	COND2	xclass
0	1	17	1	0	less than 12 hour	0	0	4	0	FD
1	1	26	0	1	less than 8 hour	0	0	3	1	MD
2	2	18	1	0	less than 8 hour	0	1	3	1	LD
3	1	19	1	0	less than 8 hour	0	1	3	1	MD
4	1	20	1	0	less than 8 hour	0	1	2	2	LD
5	2	21	1	0	less than 8 hour	1	1	2	2	LD
6	2	22	1	1	less than 8 hour	1	0	2	2	LD
7	1	23	1	0	less than 8 hour	1	0	2	2	LD
8	1	19	1	0	less than 8 hour	1	0	2	2	LD
9	1	20	0	0	less than 12 hour	0	0	2	2	FD

Figure 4.6: Visualization of Dataset

- ## **Data Error Check**

Now we have to check is there any null value available in our dataset. If there are null values then we have to remove those null values shown in Figure 4.7. But in our dataset we haven't any null value so we don't have to worry about it.

```
# look for missing values
df.isnull().any()
```

```
Sex            False
Age            False
device_use     False
feel_loss      False
time_spend     False
having_meal    False
lonely         False
COND1          False
COND2          False
xclass         False
dtype: bool
```

Figure 4.7: Null Value Check

- ## **Data Visualization**

Here we take Age and Gender Column form dataset and make it visually. From this Figure 4.8, we can easily see the number of various Ages and Gender. By seeing this figure, we can understand that, in which Ages and how much people are male or female.

```
sns.countplot(x='Age', hue='Sex', data=df, palette='colorblind', edgecolor=sns.color_palette('dark', n_colors=1))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1cc86022488>
```

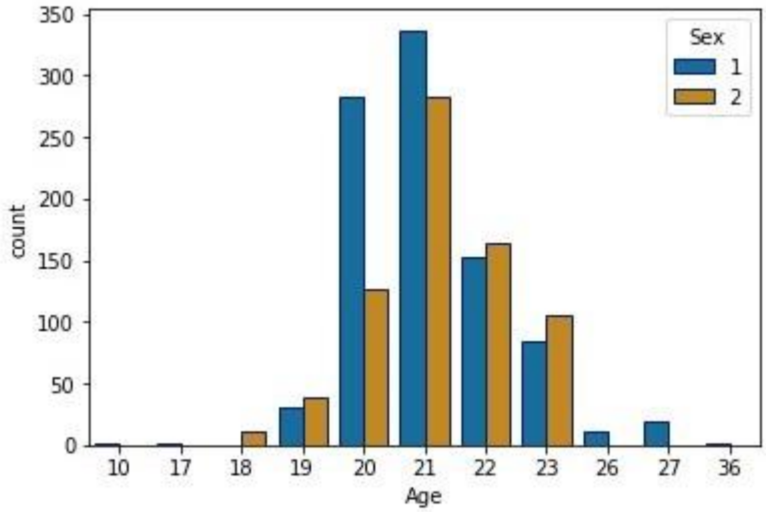

Figure 4.8: Age vs Gender

From this Figure 4.9, we can see that which age group people uses which kinds of devices. We can see most of our respondent are 20-23 year aged. They use mobile phone most.

```
sns.countplot(x='Age', hue='device_use', data=df, palette='colorblind', edgecolor=sns.color_palette('dark', n_colors=1))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1cc85b51188>
```

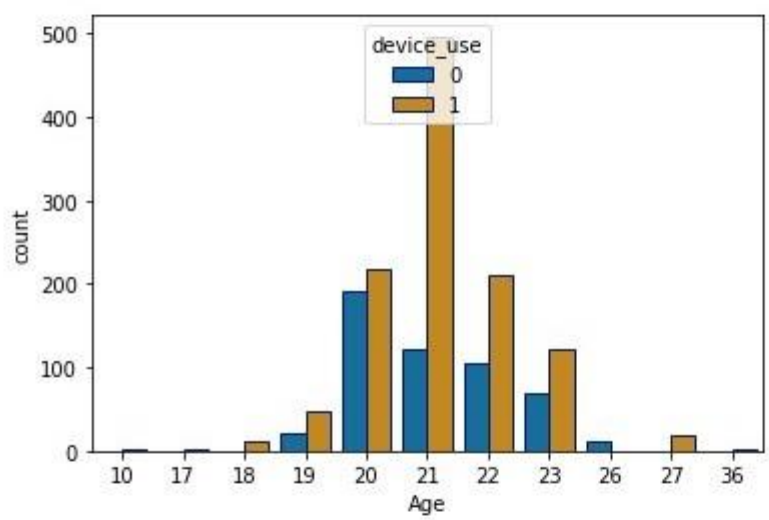

Figure 4.9: Age vs Device Use

- **Dataset after converting string data to numeric data**

From this Figure 4.10, we can see the updated numeric dataset.

	Sex	Age	device_use	feel_loss	having_meal	lonely	time_spend_numeric	class
0	1	17	1	0	0	0	1	2
1	1	26	0	1	0	0	3	1
2	2	18	1	0	0	1	3	0
3	1	19	1	0	0	1	3	1
4	1	20	1	0	0	1	3	0
5	2	21	1	0	1	1	3	0
6	2	22	1	1	1	0	3	0
7	1	23	1	0	1	0	3	0
8	1	19	1	0	1	0	3	0
9	1	20	0	0	0	0	1	2

Figure 4.10: Converted Numeric Dataset

- **Correlation analysis between variables**

Correlation analysis between all the variables shown in Figure 4.11.

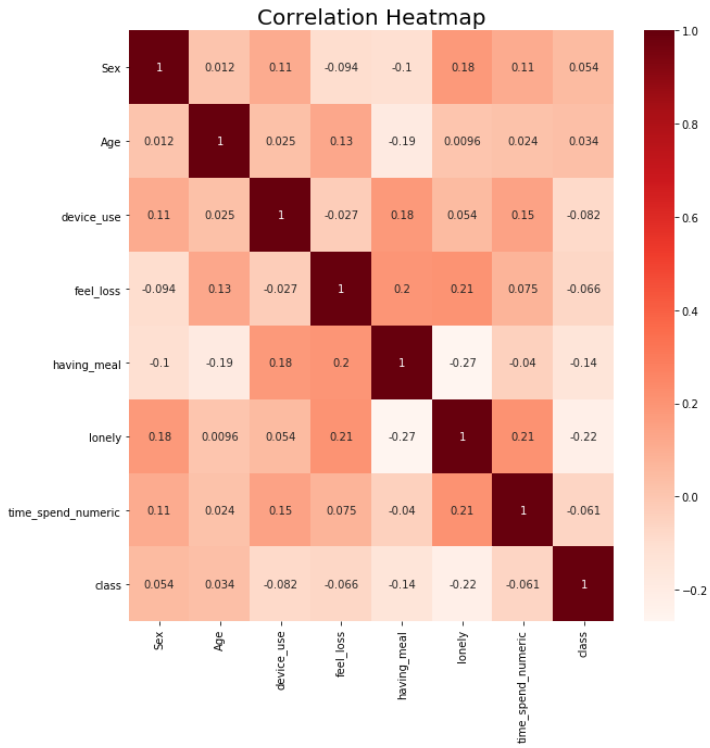

Figure 4.11: Correlation between Variables

- **Divide the data into train and test**

We divide our data into train and test split using the sklearn train_test_split function shown in Figure 4.12. We divide the data set at 80% of the training data and 20% of the test data. Test data will be selected randomly for better prediction

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20)
```

Figure 4.12: Train and Test Split of Data

4.4 Test Implementation

Finally, we use machine-learning algorithm for find out the final accuracy of our model. We have used three machine learning algorithm to complete our work which is shown in Figure 4.13, 4.14, 4.15 and 4.16.

- **Decision Tree Algorithm:**

```
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(criterion='gini', random_state=100, max_depth=None, min_samples_leaf=1)
classifier.fit(x_train, y_train)
```

Figure 4.13: Decision Tree Algorithm

- **Support Vector Machine Algorithm:**

```
from sklearn.svm import SVC
svc = SVC()
s= svc.fit(x_train, y_train)
print("SVC Model Score" , ":" , s.score(x_train, y_train) , "," ,
      "Cross Validation Score" ,":" , s.score(x_test, y_test))
```

Figure 4.14: SVM Algorithm

- **Linear Regression:**

```
# with sklearn
regr = linear_model.LinearRegression()
regr.fit(X, Y)
```

Figure 4.15: Linear Regression

- **Random Forest Algorithm:**

```
# final model
forest = RandomForestClassifier(n_estimators=36, min_samples_leaf=2)
f = forest.fit(x_train, y_train)
print("Raondom Forest Model Score" , ":" , f.score(x_train, y_train) , "," ,
      "Cross Validation Score" ,":" , f.score(x_test, y_test))
```

Figure 4.16: Random Forest Algorithm

4.5 Test Result Analysis

Mental incoherence is a significant cause of depression. In this work, we have used multiple algorithms to get more accuracy. Using the decision tree model we got 93% accuracy. From random forest and SVM algorithms we got 89% and 40% accuracy. Using the decision tree algorithm, we got our best accuracy shown in Table 4.1.

Table 4.1: Model Accuracy

Algorithm	Accuracy
Decision Tree	93%
Random Forest	89%
SVM	40%
Linear Regression	75%

4.6 Descriptive Analysis

The relation between time and other variables like behavioral changes and depression state is completely relatable. From this Figure 4.17, we notice that behavioral and stress changes are taking place when the time management changes. We named the changes as level 1 to level 5. Changes in Depression State and Behavior State are being noticed when mobile devices are being used for a long time [17]. All these changes are happening due to different hormones in our body. A treatment program was organized by the Banyan treatment center called Boca mood and anxiety disorder. They find some differences between the depressed brain and the normal brain. Depressed people have a thicker grey matter. This stage is called the gray matter abnormality. Cortisol is called the stress hormone. People with major depression disorders release cortisol in larger quantities than the average person. As a result, huge amount of exposure, parts of the brain may shrink. The amygdala is associated with emotion regulation. People with depression often have a more active amygdala than the normal brain. In particular, the amygdala in depressed people is more active than depressed people in expressing negative stimuli, such as depression [8]. There are some differences when both individual's express positive stimuli such as happy faces.

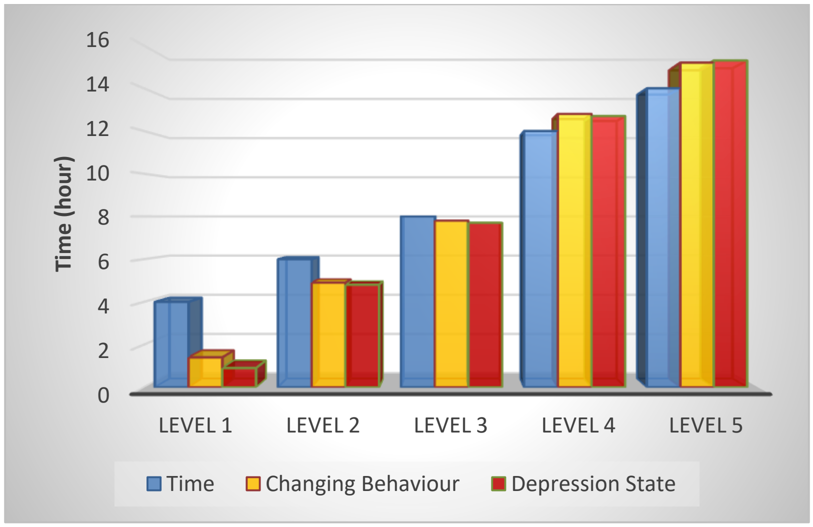

Figure 4.17: Statistics of Dependency on Device

4.5 Summery

Depression and brain both are interconnected. Excessive time spent with the device increases the amount of depression. Figure 4.18, illustrates that the green portion of this figure are those who are using or connected with device less than 1 hour to 5 hour and the yellow portion is for those who are connected with device less than 6 hours to 8 hours. The red portion of the figure for those who are more addicted to the device and use 8 hours to 12 hours. Using devices less than 1 hour to 5 hour is normal, but 6 hours to 8 hours leads to depression and those who use 8 hours to 12 hours are mentally unbalanced and depressed.

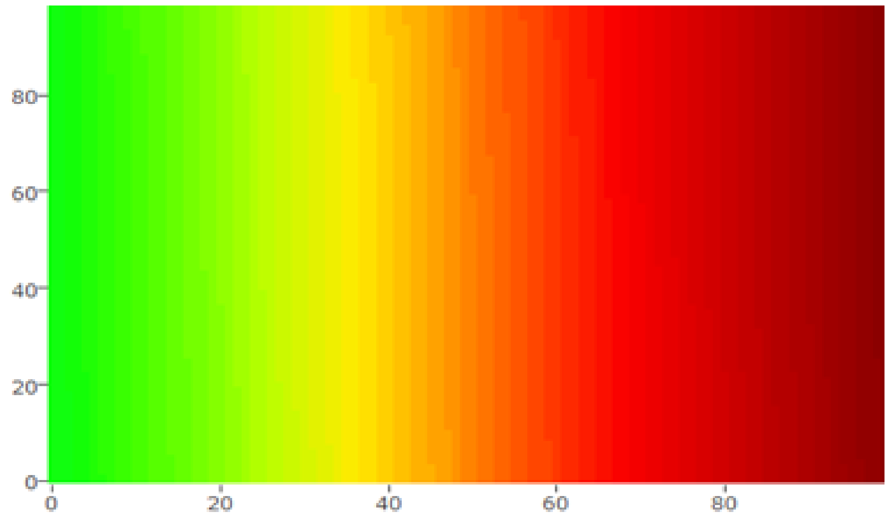

Figure 4.18: Brain Conditions Spending Too Much Time with Device

CHAPTER 5

SUMMARY, CONCLUSION AND IMPLICATION FOR FUTURE RESEARCH

5.1 Summery of the Study

Depression has become one of the major causes in human lives nowadays. Depression can cause serious changes in a person's life that can even ruin his life. There can be many causes of frustration, but the extreme use of technology is one of them. We have tried to find out how a person gets frustrated as a result of using technology in our research. With the advancement of technology, we are becoming more and more dependent on different devices day by day. This is why we are always connected to one or more devices and the use of these devices for a long time is causing a lot of changes in our brain and that is why we are getting depressed.

5.2 Conclusion

This research confirms that one of the reasons for depression is the overuse of mobile phones. The more time you spend with a mobile, the more it creates stress in the brain, which subsequently causes depression. The development of the ICT sector is looking for new horizons for us, but adolescents and young adults are becoming at risk of this new technology. Depression can be the cause of suicidal thoughts or feelings of inadequacy, personal turmoil or conflict with family, physical abuse, sleep problems, chronic pain, anxiety, etc. Since adolescents are at higher risk of depression, this is why it is more dangerous. Frustration does not allow them to move forward in life. They will be left behind.

5.3 Future Work

In this research we examined how someone can be depressed by using excessive time with device. We will briefly discuss in our future works, how to overcome this serious problem and how we can easily survive through this situation. In this work, we calculate the accuracy of the whole model but we did not calculate the level of frustration that anyone is frustrated with. In the future, we have a desire to solve this problem and provide an initial solution to depression. Provide what to do or what to avoid when someone is frustrated. Does he need to consult a doctor or not? We will create a system to do all this.

REFERENCES

[1] "American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5)", Fifth edition. 2013. National Institute of Mental Health. (Data from 2013 National Survey on Drug Use and Health.)

[2] Smith, Joanna and Firth, Jill, (2011), "Qualitative data analysis: the framework approach". Nurse Researcher, 18 (2). pp. 52-62. ISSN 1351-5578.

[3] Madhvi Saxena, Neelam Phate, K.J. Mathai, M.A. Rizvi, "Clustering Based Energy Efficient Algorithm Using Max-Heap Tree for MANET", INSPEC Accession Number: 14333798 DOI: 10.1109/CSNT.2014.33 .Publisher: IEEE.

[4] A Kumar, A Sharma, A Arora, "Anxious Depression Prediction in Real-time Social Data", Available at SSRN 3383359, 2019 - papers.ssrn.com

[5] P Cuijpers, H Riper, G Andersson, "Internet-based treatment of depression", Current Opinion in Psychology,2015 – Elsevier

[6] Farooq, R. K., Asghar, K., Kanwal, S., & Zulqernain, A. (2016), "Role of inflammatory cytokines in depression: Focus on interleukin-1β", Biomedical Reports, 6(1), 15–20. doi:10.3892/br.2016.807

[7] Kumar, S., & Chong, I. (2018), "Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States", International Journal of Environmental Research and Public Health, 15(12), 2907. doi:10.3390/ijerph15122907

[8] Kumar, A., Dogra, P., & Dabas, V. (2015), "Emotion analysis of Twitter using opinion mining", 2015 Eighth International Conference on Contemporary Computing (IC3). doi:10.1109/ic3.2015.7346694

[9] Mok, W. T., Sing, R., Jiang, X., & See, S. L. (2014), "Investigation of social media on depression". The 9th International Symposium on Chinese Spoken Language Processing.

[10] Przepiorka, A., Blachnio, A., & Cudo, A. (2019). "The role of depression, personality, and future time perspective in internet addiction in adolescents and emerging adults". Psychiatry Research, 272, 340–348.

[11] Prasenjit Saha, Prasad M.L.V. and P. Rathish Kumar , " Predicting strength of SCC using artificial neural network and multivariable regression analysis".

[12] Eisdorfer, C., Czaja, S. J., Loewenstein, D. A., Rubert, M. P., Arguelles, S., Mitrani, V. B., & Szapocznik, J. (2003). The Effect of a Family Therapy and Technology-Based Intervention on Caregiver Depression. The Gerontologist, 43(4), 521–531.

[13] Bessière, K., Kiesler, S., Kraut, R., & Boneva, B. S. (2008). EFFECTS OF INTERNET USE AND SOCIAL RESOURCES ON CHANGES IN DEPRESSION. Information, Communication & Society, 11(1), 47–70.

[14] Adams, S. K., & Kisler, T. S. (2013). Sleep Quality as a Mediator Between Technology-Related Sleep Quality, Depression, and Anxiety. Cyberpsychology, Behavior, and Social Networking, 16(1), 25–30.

[15] Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., & Tien, A. (2004). “Center for Epidemiologic Studies Depression Scale: Review and Revision (CESD and CESD-R)”. In M. E. Maruish (Ed.), The use of psychological testing for treatment planning and outcomes ass.

[16] Lenhart, A., Purcell, K., Smith, A. Zickhur, K. "Social Media & Mobile Internet Use Among Teens and Young Adults." (2010). Ed

[17] Kamal Uddin Sarker, Nazmun Nessa Moon, and Samsuddin Ahmed, “Classifying the Practitioner’s Behaviour in Medical Informatics by Using Data Mining” International Journal of Computer Science and Information Technology (IJCIT-2011), Volume-1, Issue No-2

PLAGIARISM REPORT

Predicting the Depression Level of Excessive Use of Mobile Phone: Using Machine Learning Algorithm

ORIGINALITY REPORT

9%	6%	5%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Iftakhar Mohammad Talha, Imrus Salehin, Susanta Chandra Debnath, Mohd. Saifuzzaman, Nazmun Nessa Moon, Fernaz Narin Nur. "Human Behaviour Impact to Use of Smartphones with the Python Implementation Using Naive Bayesian", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020 Publication	3%
2	www.banyanmentalhealth.com Internet Source	1%
3	ieeexplore.ieee.org Internet Source	1%
4	Imrus Salehin, Iftakhar Mohammad Talha, Md. Mehedi Hasan, Sadia Tamim Dip, Mohd. Saifuzzaman, Nazmun Nessa Moon. "An Artificial Intelligence Based Rainfall Prediction Using LSTM and Neural Network", 2020 IEEE International Women in Engineering (WIE)	1%