A MACHINE LEARNING APPROACH TO PREDICT SOCIAL MEDIA ADDICTION OF BANGLADESHI PEOPLE DURING COVID-19

BY

MEHERIN AKTER
ID: 181-15-10634
AND

KANIZ FATIMA RITU
ID: 181-15-11332

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

Supervised By

MD. TAREK HABIB
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2022

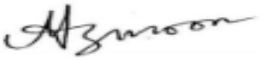
APPROVAL

This Project titled **"A Machine Learning Approach to Predict Social Media Addiction of Bangladeshi People during COVID-19"**, submitted by **Meherin Akter**, ID No: **181-15-10634** and **Kaniz Fatima Ritu**, ID No: **181-15-11332** 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 4th January, 2022.

BOARD OF EXAMINERS

Chairman

Dr. Md. Ismail Jabiullah
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Nazmun Nessa Moon (NNM)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Aniruddha Rakshit (AR)
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Md. Arshad Ali
Associate Professor
Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Md. Tarek Habib, 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:

Md. Tarek Habib
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Meherin Akter
ID: 181-15-10634
Department of CSE
Daffodil International University

Kaniz Fatima Ritu
ID: 181-15-11332
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project successfully.

We really grateful and wish our profound our indebtedness to **Md. Tarek Habib**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Research in Machine Learning influenced us 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 project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan**, Head, Department of CSE and **Md. Tarek Habib**, Assistant Professor, Department of CSE for their 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 discussion while completing the course work.

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

ABSTRACT

The world is in an extremely precarious situation, with coronavirus posing a serious threat. To be safe, staying at home is the best option right now. People nowadays spend the significant amount of their time on social media platforms. Just as social media has stood by people during this pandemic, it has also caused trouble in some cases. Excessive use of social media has an adverse effect on mental and physical wellbeing. Besides, the over usage of social media leads to negative impact on once social life, moreover it can also lead people to make various dangerous crimes. In this research study, the use of social media by Bangladeshi people throughout the year 2021 was examined in order to anticipate their level of addiction to this COVID-19 circumstance. The data was gathered from people of various occupations, and the levels of addiction has been analyzed with the help of professionals and several internet sites. Using some methods and machine learning classifiers, their addiction to social media has been predicted in which the levels are categorized into four class labels. Several data mining techniques, methods and machine learning classifiers have been applied and found the maximum accuracy, 94.05% in logistic regression.

CONTENTS **PAGE**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

In today's world, the internet is a need. The Internet has turned our existence upside down by revolutionizing communications, to the extent that it is now our preferred medium of everyday communication and in almost everything we do, we use the Internet. Ordering a pizza, buying a television, sharing a moment with a friend, sending a picture over instant messaging [1]. Many national and international organizations use the internet to provide a variety of services all around the world. Social media has become a leading entertainment source to many people particularly after the pandemic appeared. The term social media refers to a computer-based technology that facilitates the sharing of ideas, thoughts, and information through virtual networks and communities and it is internet-based and gives users quick electronic communication of content, such as personal information, documents, videos, and photos [2]. An analysis shows that there are 4.55 billion social media users around the world in October 2021, equating to 57.6 percent of the total global population [3]. Since early March 2020, the new Coronavirus Disease (COVID-19), which the World Health Organization has designated a pandemic, has spread to Bangladesh, with people receiving daily updates via social and electronic media [4]. Social networking services such as Facebook, Instagram, and Twitter have made people's life easier in the twenty-first century. They, on the other hand, hold a different viewpoint. Social media addiction has developed to the point that it is causing havoc on the psyche of individuals who are immersed in the virtual world. Now a days the most used social media platforms are Instagram, WhatsApp, Facebook, Messenger, Snap chat, TikTok and Twitter. Excessive usage of social media is harmful for everyone and lead to addiction to social media. Social media addiction is a behavioral addiction that is characterized as being overly concerned about social media, driven by an uncontrollable urge to log onto or use social media, and devoting so much time and effort to social

media that it impairs other important life areas [5]. Mainly young people are addicted to social media. The addiction of social media vastly increases all around the world in the pandemic situation of covid-19. As a developing country like Bangladesh, people are also addicted to social media due to covid-19. The addiction of social media is harmful for the society. Social media addiction also causes mental problems such as frustration, irritation, anxiety of physical communication etc. [6]. Again, social crimes also raises due to this addiction. So in order to avoid social media addiction, we need to use social media in a proper way. In this study, we've tried to analyze the addiction level of Bangladeshi people in this crisis situation using machine learning. We've analyzed the collected data under four classes. Since Logistic Regression did best in term of accuracy, we've implemented the predictive model. This research will help every individual to understand their addiction level and also help to take necessary steps to make themselves positive in this situation of pandemic.

1.2 Motivation

Due to the pandemic the world becomes more dependent on social media. All the work done from home by internet. School, college, university, office, court etc. work done by any of the social media platform. Some people do some vulnerable crime on social media such as harassment, hate speech, child abusing, inappropriate comments, sexual content etc. [5]. Most of the people in Bangladesh do not know how to use social media and knowingly or unknowingly do such crimes. In 2020, Bangladeshi all-rounder cricketer became the victim of cyberbullying because some Facebook users posted offensive and child abusing comments on his 4 year old child [7]. The tendency of young people now is to increase followers and become famous on social media, for this reason many young people do unethical content, follow western culture and forget their own culture. Objectively, social media are not bad but the problem with their popularity among children is they increase the risk of being exploited by those involved in human

trafficking and sex crimes. [8]. Government and parents should monitor and focus on social media that none can addicted to social media. So our motivation behind this research is awaking the people of our country by figuring out the addiction level toward social media. We use machine learning for predicting the addiction of social media due to covid-19.

1.3 Rationale of the Study

The discovery of Social Media Addiction at the right time can help us to stop wasting time. It is not easy to find out that anyone is addicted to social media at the beginning. For this reason, Machine Learning approach is a massive help for identifying addiction. We can reduce the rate of addiction if we can detect it early on using various data mining techniques.

1.4 Research Questions

- How we'll identify the addiction of social media?
- How we'll collect the data set during pandemic?
- How will our original data set look like?
- Do we need to train our original data for machine learning approach?
- How much data we need to collect?
- Will our data set and machine learning be compatible?
- Which technique of machine learning we should use?
- How we'll aware the people about the addiction of social media?

1.5 Expected Outcome

The people of Bangladesh are not concerned about the usage of social media. So they do not easily find out how come they addict to social media and it is really hard to find out the addiction from the beginning. The addiction of social media is extremely harmful. A person who is addicted is harmful for the society and we need to stay away from the addicted person. There is none other than awareness can only decrease the addiction of social media. From our research we can only aware people and change their life style by showing the predicted risk of excessive usage of social media. Some research has already done about social media addiction and the risk of the prediction is based on a little numbers of features that's why the accuracy rate of risk of Social Media Addiction is not enough but in our research we analyze a lot of features like age, gender, spend time, status update anxiety, compare with real life, excessive time, add new friends, Facebook design, share personal life, check Facebook, share image, spending hours browsing, virtually date, fake id, Facebook Addiction, change mood, easiest way to contact, part of an exciting world, Facebook games, selfie post, food page, religious views, feel urge to use Facebook, forget personal problem etc. as a result we are able to show the risk of addiction with higher accuracy. Besides, the building of a data set for social media addiction in the context of Bangladesh. Publication of one or more articles in international conference proceedings or journals.

1.6 Project Management and Finance

When we started our research, we had a very little knowledge about machine learning and research procedures. Our honorable supervisor helped us a lot throughout the whole process. The data collection process was done by me, Kaniz Fatima Ritu and the implementation process was done by my team mate, Meherin Akter. And in our report,

Chapter 1, 5 and 6 was written by me, Kaniz Fatima Ritu and Chapter 2, 3 and 4 were written by my mate, Meherin Akter.

As we started our research during lockdown, so first we collected data via electric form which didn't need any financial help. But when the pandemic situation was a bit normal then we printed the survey form which costs some money and we spent it from our self-fund.

1.7 Report Layout

This research paper contains the following contents as given below:

- In Chapter one we discuss about the introduction of the research let alone its motivation, rationale of the study, research questions, and expected outcome.
- In Chapter two we discusses about the related works, research summary, the scope of the problem, and challenges.
- In Chapter three we discuss about the workflow of this research, data collection procedure, and statistical analysis and feature implementation.
- In Chapter four we discuss experimental evaluation and some relevant discussions, the outcome of research via numerically and graphically.
- In Chapter five we cover this research impact on society.
- Chapter six contains a summary of this research work along with the limitation and future work.

CHAPTER 2

Background Study

2.1 Introduction

We examine relevant works, a study synopsis, the breadth of the topic, and problems in this part. We outline several research papers, related works, underlying methodologies, classifiers, and accuracies that are relevant to our study in the related work section. We produce a summary of selected relevant studies and put them in a table for better and easier understanding in the research summary section. The breadth of the problem section outlines how our work paradigm can contribute to the problem. Finally, in the Challenges section, we discuss the hurdles and dangers we encountered while conducting this research.

2.2 Related Works

This research paper's literature review portion will show recent related works on social media and addiction prediction done by several researchers. We have followed and analyzed their work in order to gain a better understanding of the procedures and strategies they used in their research.

Sabbir Alam et al. [15] analyzed the emotion of Bangladeshi people during Covid-19 in social media. The opinion of Bangladeshi people about this pandemic was determined by analyzing their comments on many Facebook news posts linked to coronavirus. Deep learning has been used at this research. A research into their emotions was conducted using three classes (Analytical, Depressed, and Angry). The Bangla language was used to create the data set. Several deep learning algorithms were used, with the highest accuracy being obtained in CNN (97.24%) and LSTM (95.33%). The majority of people

commented analytically, as evidenced by the results. The findings reveal Bangladesh's public perception of the pandemic.

Wahyu Rahardjo and Indah Mulyani [16] researched on the Instagram addiction among teenagers. The goal of their study was to perceive how type D personality (a shared proclivity towards negative affectivity), self-esteem, and fear of missing out (FoMO) affect the teenagers who are addicted to Instagram. A total of 259 people took part in the study, all of whom are Type D personalities with a serious Instagram addiction who live in Jakarta. A larger area was obtained as a result of the screening results (based on their total scores). For data gathering and analysis, this study used snowball sampling and regression approaches. According to the findings, all factors influenced 56.9% of Instagram addiction cases.

Johura Khatun et al. [17] have used machine learning for detecting betel nut addiction, in their research. K-nearest neighbor (kNN), Support Vector Machine (SVM), decision tree, logistic regression, random forest, and Naïve Bayes algorithm have been used [1]. The random forest has achieved 99.00% accuracy with less training time and KNN hasn't shown well performance in their research, achieved 95.52% which was the lowest accuracy. They have used eighteen features in their research, 800 records have been used for training and 201 records have been used for testing. Precision, recall, accuracy, and F1-score have been used in their study to evaluate the performance of machine learning algorithms.

Zahirul et al. [18] devised a method for predicting the likelihood of being addicted to Facebook using machine-learning algorithms. In their study, they identified people's addiction to Facebook as well as their knowledge of their daily routine. In total 8 machine learning algorithms such as support vector machine, decision tree, naïve Bayes, linear regression, logistic regression, linear discriminant analysis, k-nearest neighbors, and

random forest were used to predict the result of addiction. They evaluated each of these classifiers' performance using a variety of well-known performance criteria. They employed principal component analysis to mathematically reduce data. Their findings show that SVM surpasses the other algorithms by having a higher accuracy rate which was 85%.

CIPLAK and Ersun [19] examined social media addiction on Turkish people using variable prediction. Average weekly time spend on social media, happiness and narcissism were the tested hypothesis of their study. The study group comprised of 239 people (140 women (58.6%) and 99 men (41.4%)), all of whom lived in different parts of Turkey. The Social Media Addiction Scale, the Scale of Happiness - Short Form, the Narcissistic Admiration and Rivalry Scale, and the Personal Information Form were used to collect data for this study. Frequency, percentage, independent sample t-test, Pearson correlation coefficient, and stepwise regression analysis were calculated to examine the data. In accordance with the findings of this study, social media addiction differed noticeably from the average weekly time spent on social media. According to the findings, both narcissism and average time spent on social media positively predicted social media addiction whereas happiness represented it negatively.

Md. Rashedul Hasan [20] conducted a study to determine the extent of social network addiction in the real lives of young people in Bangladesh during COVID-19. In order to complete his study, he conducted an online interview. In this poll, 520 young people from various universities in Bangladesh were interviewed. The findings suggest that many people are addicted to social media sites, and that they are also experiencing mental and physical health issues during COVID-19. . SPSS v26 was used to analyze his collected data.

In their study, Aydin et al. [21] investigated at the impacts of social media addiction on adult depression. They wanted to see if different factors (age, gender, time spent on social media each day, highest level of education, frequency of use, etc.) had distinct effects on social media reliance. 419 individuals from diverse Turkish regions, ranging in age from 18 to 62 years old, participated in the study. The Social Media Dependence Scale (SMDS), Beck Depression Inventory scores, and demographic information were collected using the questionnaire form. The research was conducted using a typical screening model. There were substantial differences between depression and reliance of social media based on factors such as age, the number of children and income. When social media addiction was studied in terms of gender among socio-demographic characteristics as a consequence of the study, no significant difference was identified.

İmran and Mehmet [22] measured social media addiction among university students of Turkey. The goal of their study was to find out how social media is used by the university students and what level of SMA they have according on their age, gender, usage years, and daily usage. To reach this goal, in 2019, they questioned 665 students at Bingöl University via an online platform. The SPSS 20 program used descriptive statistics, the Kruskal Wallis Test, One-Way ANOVA test, Factor Analysis, and Correlation Analysis statistical methods to analyze and validated data. Instagram, YouTube, and Facebook revealed to be the three most popular social media networks respectively. The element of daily consumption had the strongest link with the component of occupation in their study. Furthermore, there is a strong correlation between relapse and conflict variables, with the confrontation factor on the Social Media Addiction Scale being the most indispensable ingredient (SMAS).

Haydar Hoşgöret et al. [23] have used IBM SPSS V.22.0, t test, ANOVA, Pearson's correlation, and simple regression analysis for investigated the relationship between addiction to social media and job among nurses in Turkey. The confidence interval was

approved for 95% in the evaluation of the findings derived from their study, and the importance was determined to $p < 05$. The result of social media addiction is 2.42 ± 0.95 and work engagement is 4.28 ± 1.20 respectively. 205 participants were examined, of which 48 were male and 197 were female.

Rahmatullah HAAND and Ahmad Zia ELHAM [24] have used probability sampling method for their research to find out the addiction of social media of university students in Afghanistan. They found that 5.95% of precipitants were addicted seriously. 384 participants were examined, of which 361 were male and only 23 were female, aged between 18-25 years in the research.

Mustafa USLU [25] has used IBM SPSS Statistics 25 program for his research to find out the addiction of social media usage of Turkey people. He observed that people are not uneasy with social media, but satisfied. The study also found that Instagram is the most common application of social media with a proportion of 89.66%. The addiction of Social media among 14-18 years of age and unemployed was the highest. He collected the data via internet.

2.3 Comparative Analysis and Summary

There are some work has already done about prediction and detection with the machine learning algorithm and data mining process. Nowadays, the use of machine learning technology has increased with the use of alcohol user prediction, tobacco user detection, and various disease detection. The comparison between these related works has shown in this part. Here, the comparison of different research works with their subject, methodology, and the outcome are given below in Table 2.1.

TABLE 2.1: SUMMARY OF RELATED RESEARCH WORK.

SL	Authors	Objects(deal with)	Problem Domain	Size of feature sets	Algorithm/Method	Accuracy
1.	Sabbir Alam et al.	Emotion Analysis of Bangladeshi People during Covid-19 in Social Media	Prediction	1120	CNN	97.24%
2.	Wahyu Rahardjo and Indah Mulyani	Instagram addiction among teenagers.	Perception	434(total) 259(D-type)	Snowball sampling and regression techniques	56.9% (affected)
3.	Johura Khatun et al.	Betel nut addiction	Detection	1001	Random Forest	99.00%
4.	Zahirul et al.	Facebook Addiction Detection	Detection	1001	SVM	85%
5.	CIPLAK and Ersun	Social Media Addiction: Time spent, Happiness, Narcissism	Analysis	239	Frequency, percentage, independent sample t-test, Pearson correlation coefficient, and stepwise regression	Time spent, Narcissism: Positive Happiness: Negative

6.	Md. Rashedul Hasan	Social Media Addiction on Young People during COVID 19	Analysis	520	SPSS v26	NM
7.	Aydin et al.	Social Media Addiction and Depression among university students of Turkey	Analysis	419	Typical Screening Model	NM
8.	İmran and Mehmet	Social Media Addiction	Analysis	665	Descriptive Statistics,	8.
9.	Haydar Hoşgöret et al.	Relationship between addiction to social media and job among nurses	Analysis	205	IBM SPSS V.22.0, t test, ANOVA, Pearson's correlation, and simple regression analysis	NM
10.	Rahmatullah HAAND and Ahmad Zia ELHAM	Addiction of social media of university students in Afghanistan	Analysis	384	Probability sampling method	NM
11.	Mustafa USLU	Addiction of social media usage among Turkish people	Analysis	5176	Descriptive study in survey model	NM

Currently, new technologies that combine machine learning, artificial intelligence, and deep learning are being investigated for use in any type of prediction and detection model. Various machine learning techniques have lately been used to diagnose, predict and detect content. For any detection model, ANN, kNN, CNN, SVM, logistic regression, and other methods are common. According to prior research, the popularity and effectiveness of the kNN, SVM, random forest, ANN, Naïve Bayes, and Decision tree algorithms for prediction or detection models are high. In this study, we used kNN, SVM, logistic regression, naïve Bayes, random forest, and other algorithms to predict the probability of being addicted to social media in Bangladesh.

2.4 Scope of the Problem

The research work of ours is mainly building a model by analyzing data and applying machine-learning algorithms. Our proposed model can predict the risk of becoming addicted to social media. This prediction will have a significant impact on society. The young generation can be aware about the social media usage. Interested people who are not willing to consult with psychiatrist will be able to use this model to learn about their addiction state to social media. It is dangerous to become addicted to social media because nowadays it is very common to commit crime using social media platforms, and people specially our young generations can be destroyed. Therefore, this model would be useful for ordinary people and conscious people to keep track about their addiction to social media. Parents worry too much about their children as most the teenagers use social media which could hinder the development of a generation and reflect its adverse effects on society unless they were aware of the safe use of social media. And our research will be definitely a help in this regard. Recently, as machine learning and artificial intelligence are being used for various object detection and disease predictions, the results are quite acceptable. Therefore, we decided that using machine learning, we would create a model of addiction risk prediction.

2.5 Challenges

While doing our research we encountered some problems. Data collection was a bit challenging for us because we were not getting enough responses via the google form. For that reason we had to collect data talking to people in person which was a difficult task in this pandemic situation. Besides, ordinary people and social media addicts cannot be easily distinguished. We read a lot of newspapers, articles and talked to different people but most of the people was not going to give any information about their social media usage. It was very difficult to collect proper information because we didn't know many experts who can justify our target class. After that we had been able to collect our data from the students of different schools, colleges and universities and different professions as well. We searched for some more mental hospitals but they did not want to help us with any information because the hospital authorities had privacy issues. Finally we had been able to talk with several experts which was really a difficult task as they were very busy with their patients.

Anaconda, Jupyter Notebook, and various new machine learning algorithms were also unfamiliar to us. It took us a while to understand and learn about it at first, but with the support of our supervisor and additional practice, we were able to grasp it quickly. Then we continue to do our jobs properly and enthusiastically.

CHAPTER 3

Research Methodology

3.1 Introduction

The purpose of this research is to establish a model for predicting the addiction level to social media. The Prediction Model is created based on the daily life information of people and some other related information. To create this model, we have applied various machine-learning algorithms. We have basically used kNN, Logistic Regression, SVM, Naïve Bayes, Decision Tree, XGboost classifier and Random Forest in this research. Algorithms have been used in the model for classification purposes. We used seventeen key factors that were very closely connected with addiction. We analyzed some of the features that were responsible for the outcome. We processed our dataset before implantation. We have used six methods and compared them. We calculated and computed the accuracy of each algorithm to select the appropriate algorithm for the model.

3.2 Data Collection Process

This section was very difficult for us as it is the main stage for our research, so it was our main concern to collect the accurate data, but people didn't want to interact with us because of the COVID-19 epidemic. But somehow we managed to collect data in various ways. Therefore, we created the form and provided it to the people in neighborhood also created google form and post them into different online groups and we collected our data based on the following factors:

1. Your age?
2. Your gender?
3. From which district do you belong to?

4. Your marital status?

5. Your occupation?

6. If you are a student then what is your current education level?

7. At an average how many hours do you spend on social media every day?

8. What is your most favorite social media platform?

9. How much time do you spend on your favorite social media?

10. Your addiction level in virtual world become higher after the Covid-19 appear. Is it true?

11. When you are depressed, stressed or frustrated, social media comforts you. Do you agree?

12. What is the most likely reason of using social media?

13. When you are in trouble, do you think can help you out?

14. Which time of the day you use social media the most?

15. If you do not use social media a day or a couple of days then do you feel bad?

16. Does using social media leave a negative impact on your sleep?

17. In your leisure time would you like to prefer using social media?

18. Would you like to share your feelings at social media?

19. Do you prefer using social media over socializing offline?

20. Do you get annoyed when anybody or anything interrupts you when using social media?

To identify the risk of becoming addicted to social media we had to consider each of these factors. We find out about these factors by talking to several professionals and with the help of some websites and articles [10] [11] [12] [13].

3.3 Research Instrumentation

Predictive models can help people to make prediction for the upcoming situation. Data mining, machine learning algorithms and deep learning are very popular for any kind of prediction and detection. We have applied various algorithms to our collected data to see which algorithm will perform well for our model. We have used various machine-learning algorithms like *k*-NN, logistic regression, support vector machine (SVM), naïve Bayes, decision tree, random forest and XGboost. We have used 'Python' as a programming language and 'Anaconda navigator', 'Jupyter notebook' as a data mining tool and 'Microsoft Excel' as our dataset in our research work.

3.3.1 Proposed Methodology

At first we have collected data from 504 people using survey form, and then we processed our dataset. After preprocessing of our dataset, we used seven feature selection methods such as taking all features, principle component analysis (PCA) with 85%, 90% and 95% variance, extremely randomized trees classifier (extra trees classifier) and univariate selection. In all methods, we have used seven machine learning classifiers which are logistic regression (LR), Naive Bayes (NB), XGBoost, support vector machine (SVM), *k*-nearest neighbors algorithm (*k*-NN), random forest (RF), and decision tree (DT). Lastly, we evaluated the performance of each method and classifier. Based on performance, we selected the best method and classifier.

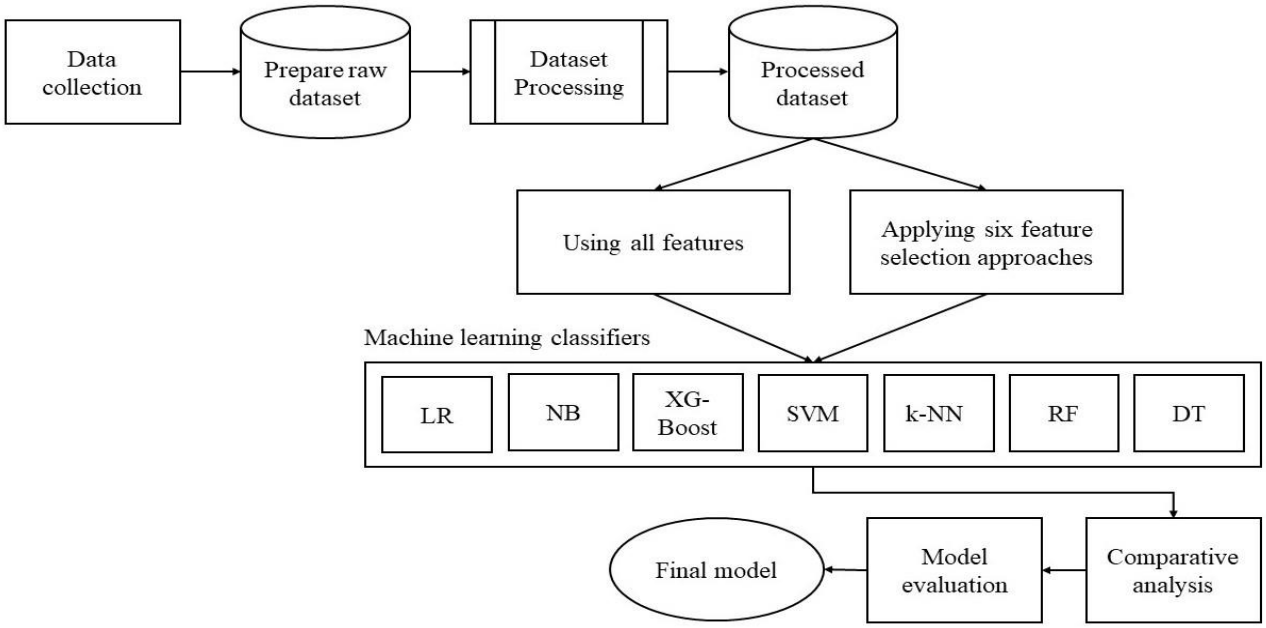


Figure 3.3.1: Proposed methodology of the research

3.3.2 Data Preprocessing

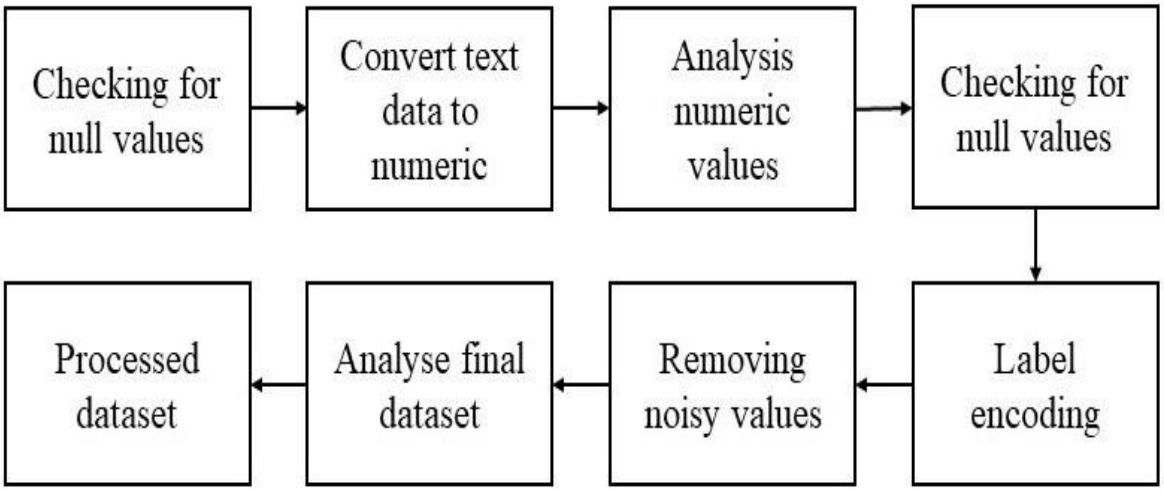


Figure 3.3.2: Data preprocessing steps

Data preprocessing is the first and most important stage in developing a machine learning model which involves preparing raw data and making it acceptable for a machine learning model because real-world data sometimes contains noise, missing values, and is in an unsuitable format that cannot be used directly in machine learning models, so data preprocessing is a necessary task for cleaning data and making it suitable for a machine learning model, which improves the model's accuracy and efficiency [34]. We acquired some categorical data, numerical data, and text data from the collected data. Then we decided to use data processing to make this data suitable for algorithms. First and foremost, begun the data cleansing process. If the data set has a null value, we proceeded to the next step. The level that turns text data to numerical data has been encoded. We were able to find the missing value issue. We then used a box plot to see if the data set had any noisy values. There was some noisy data in the numerical data. After that, we looked at the correlation heatmap as a method of data integration. The ratio of each data connected to each data is shown in this map. We removed noisy values by using outlier quantile detection. Then we dropped our outcome features that was the district column. Separate histogram of each feature helped us with data reduction and data visualization in feature engineering. Through normalization, we completed the data transformation. This whole process of data processing was done by using the "Jupyter Notebook" and "Anaconda navigator". The correlation heatmap is shown in Figure 3.3.3.

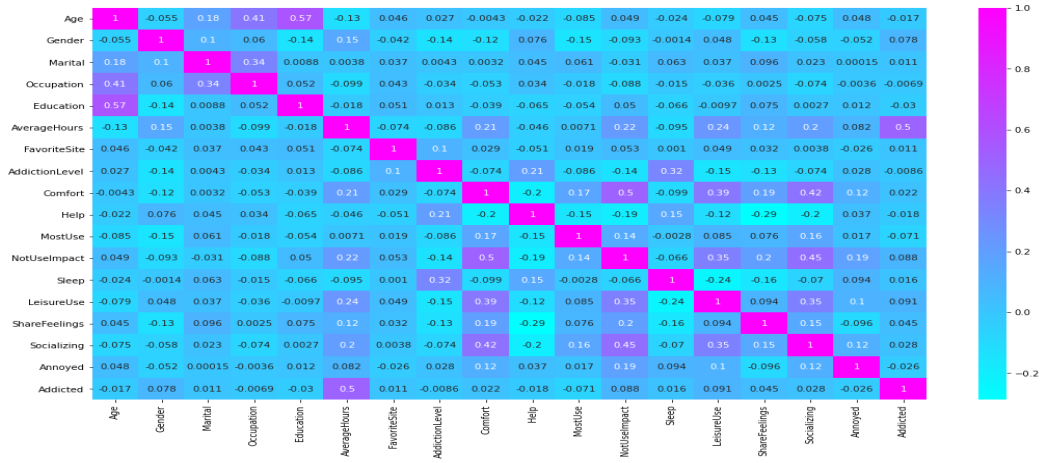


Figure 3.3.3: Correlation heatmap

3.4 Statistical Analysis

We created eighteen features based on questions of survey form, and questions and features name are given below in table 3.4.

TABLE 3.4: LIST OF FIGURE NAME ACCORDING TO SURVEY QUESTIONS

SL	Questions	Feature Name
1	Your age?	Age
2	Your gender?	Gender
3	Your marital status?	Marital
4	Your occupation?	Occupation
5	If you are a student then what is your current education level?	Education
6	At an average how many hours do you spend on social media every day?	AverageHours
7	What is your most favorite social media platform?	FavouriteSite
8	Your addiction level in virtual world become higher after the Covid-19 appear. Is it true?	AddictionLevel
9	When you are depressed, stressed or frustrated does social media comfort you?	Comfort
10	When you are in trouble do you think social media can help you out?	Help

11	Which time of the day you use social media the most?	MostUse
12	If you do not use social media a day or a couple of days then do you feel bad?	NotUseImpact
13	Does using social media leave a negative impact on your sleep?	Sleep
14	In your leisure time would you like to prefer using social media?	LeisureUse
15	Would you like to share your feelings at social media?	ShareFeelings
16	Do you prefer using social media over socializing offline?	Socializing
17	Do you get annoyed when anybody or anything interrupts you when using social media?	Annoyed
18	Your addiction level is-	Addiction

3.4.1 Conversion of Text Feature Value to Numeric Value

TABLE 3.4.1: DESCRIPTION TEXT FEATURE VALUES INTO NUMERIC VALUES

SL	Feature name	Value name	Description	Figure Number
1	Age	1	Age Range: 13-19	1
		2	Age Range: 20-30	

		3	Age Range: 31-40	
		4	Age Range: 41-60	
2	Gender	1	Male	2
		2	Female	
3	Marital	1	Single	3
		2	Married	
		3	Other	
		4	Divorced	
4	Occupation	1	Student	4
		2	Unemployed, Housewife	
		3	Other profession	
5	Education	1	School	5
		2	College	
		3	University	
		4	Other	
6	AverageHours	1	Time Spend: <2 hours	6
		2	Time Spend: 2-5 hours	
		3	Time Spend: 5-8 hours	

		4	Time Spend: >8 hours	
7	FavoriteSite	1	Facebook	7
		2	Instagram	
		3	YouTube	
		4	WhatsApp	
		5	Snapchat	
		6	Twitter	
		7	Other	
8	AddictionLevel	1	Yes	8
		2	No	
		3	Maybe	
9	Comfort	1	Strongly disagree to strongly agree	9
		2		
		3		
		4		
		5		
10	Help	1	Yes	10
		2	No	
		3	Sometimes	
11	MostUse	1	Late night	11
		2	Evening	

		3	Afternoon	
		4	Noon	
		5	Early morning	
		6	Evening; Late night	
		7	Early morning; Late night	
		8	Afternoon; Evening; Late night	
12	NotUseImpact	1	Strongly disagree to strongly agree	12
		2		
		3		
		4		
		5		
13	Sleep	1	Yes	13
		2	No	
		3	Maybe	
14	LeisureUse	1	Strongly disagree to strongly agree	14
		2		
		3		
		4		
		5		

15	ShareFeelings	1	Sometimes	15
		2	No	
		3	Yes	
		4	Most of the time	
16	Socializing	1	Strongly disagree to strongly agree	16
		2		
		3		
		4		
		5		
17	Annoyed	1	No	17
		2	Yes	
		3	Maybe	
18	Addiction	0	No addiction	18
		1	General addiction	
		2	Moderate addiction	
		3	Severe addiction	

3.4.2 Statistical Analysis of Age Range

In our dataset, there are four different age ranges which are 1. 13-19, 2. 20-13, 3. 31-40 and 4. 41-60. Among 504 participants almost 89% people were under the 2nd range which

was the highest. Almost 8%, 3% and 1% people were under the 1st, 3rd, and 4th age ranges respectively. The histogram of age analysis are shown in figure 3.4.2.

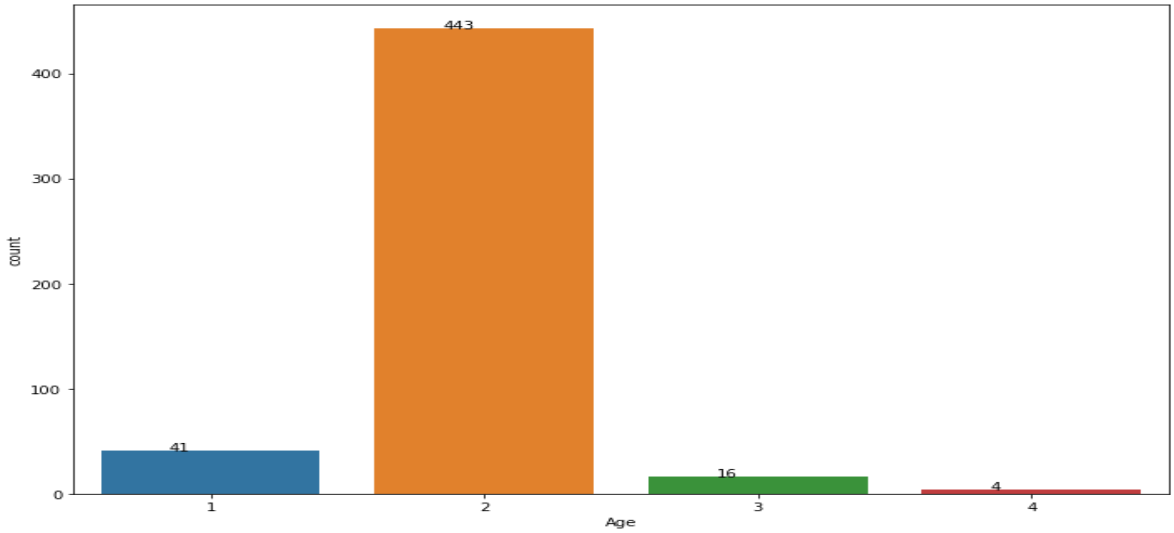


Figure 3.4.2: Histogram of participants' age range

3.4.3 Statistical Analysis of Gender

Among 504 participants of our study, almost 60% were male and 40% were female. The statistics of gender is shown in Figure 3.4.3.

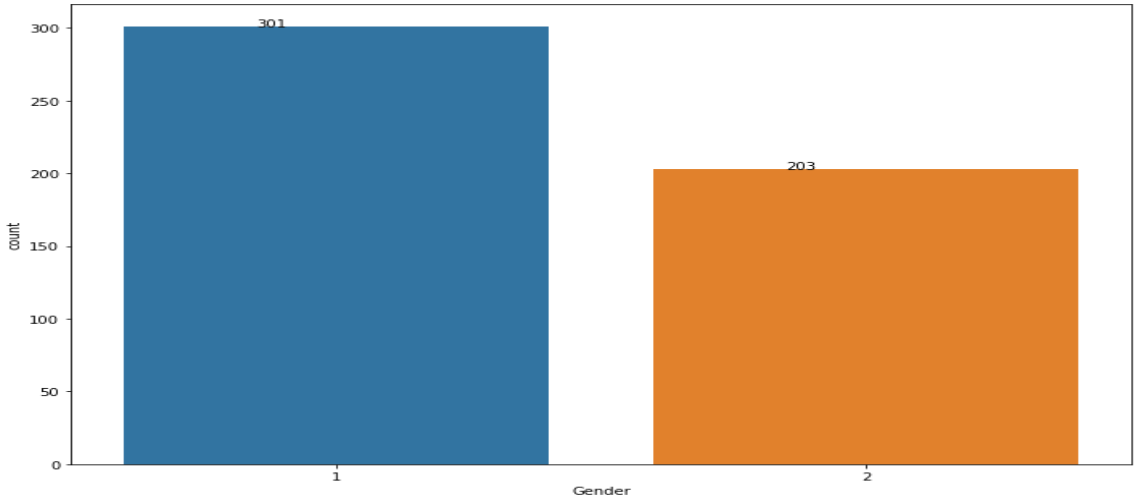


Figure 3.4.3: Histogram of gender

3.4.4 Statistical Analysis of Other Features

Other features such as marital status, occupation, education, average hour of spending time, favorite site, addiction level due to COVID-19, comfort using social media, most use time, if social media is a help, not use impact, social media usage impact on sleep, using social media in leisure time, sharing feelings in social media, socializing offline or social media and annoyed feeling are shown statistically in Figure 3.4.4 to Figure 3.4.18.

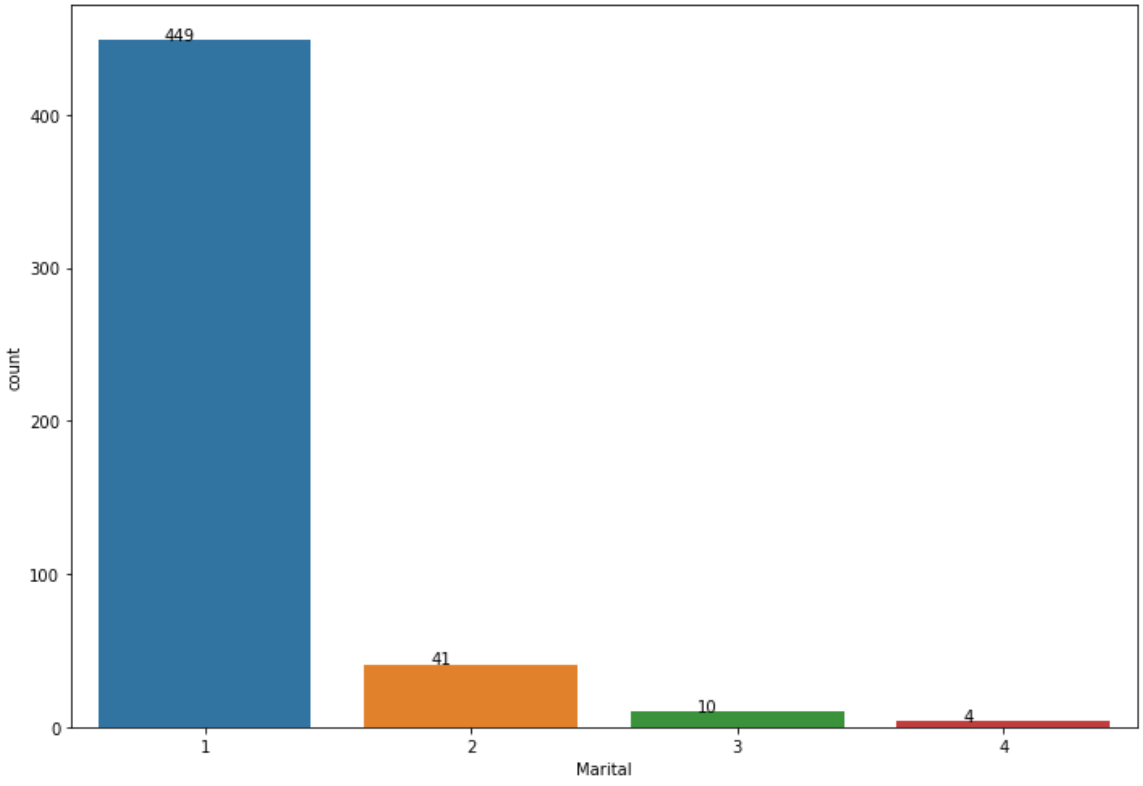


Figure 3.4.4: Histogram of marital status

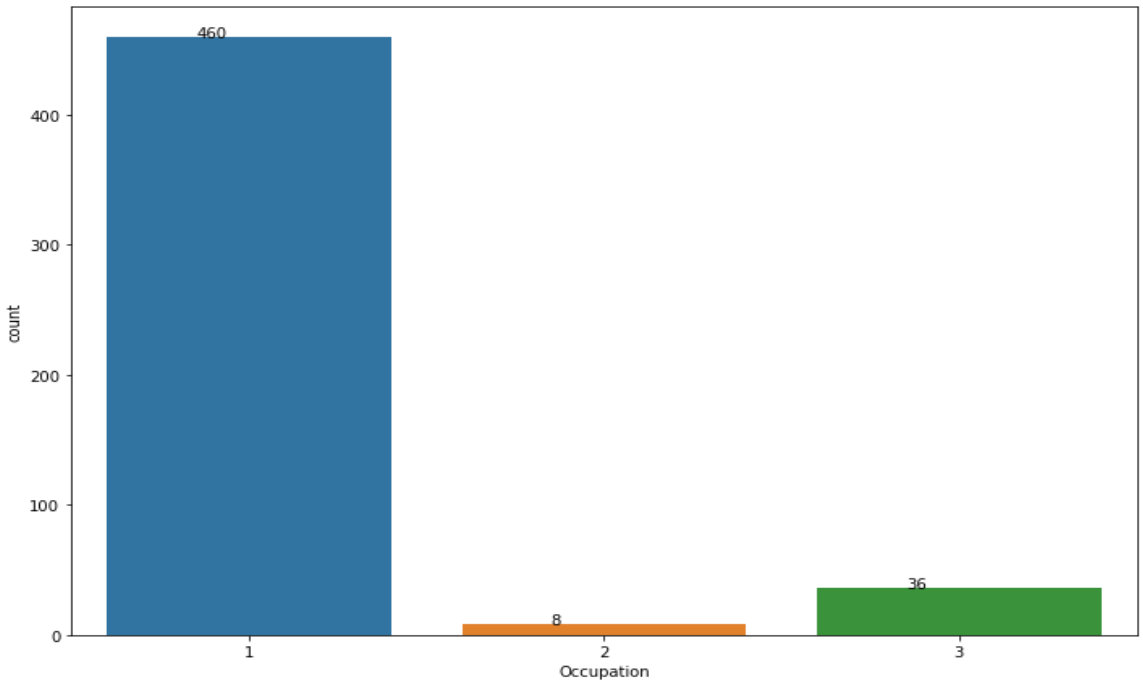


Figure 3.4.5: Histogram of occupation

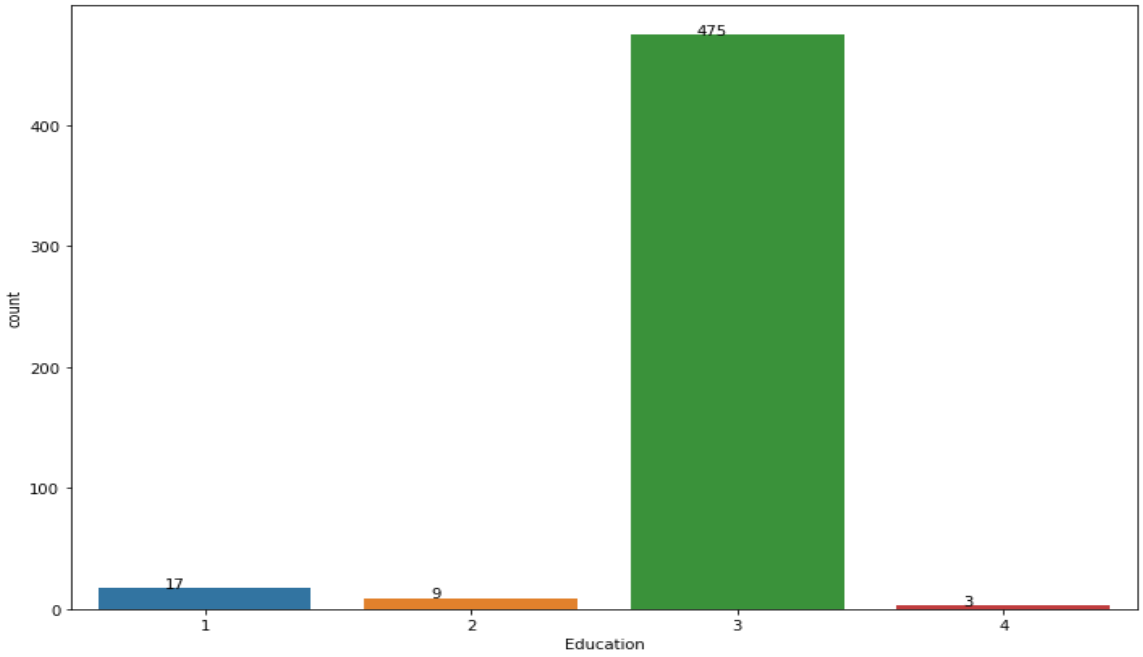


Figure 3.4.6: Histogram of education

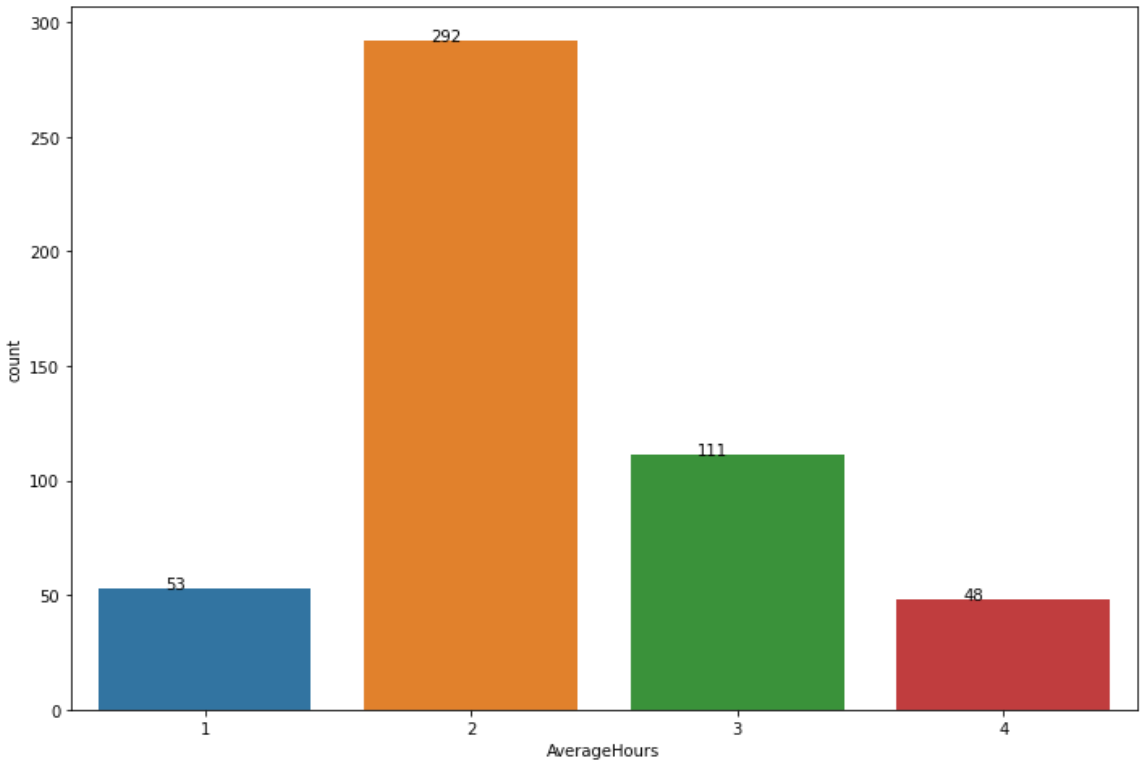


Figure 3.4.7: Histogram of average hour of time spend

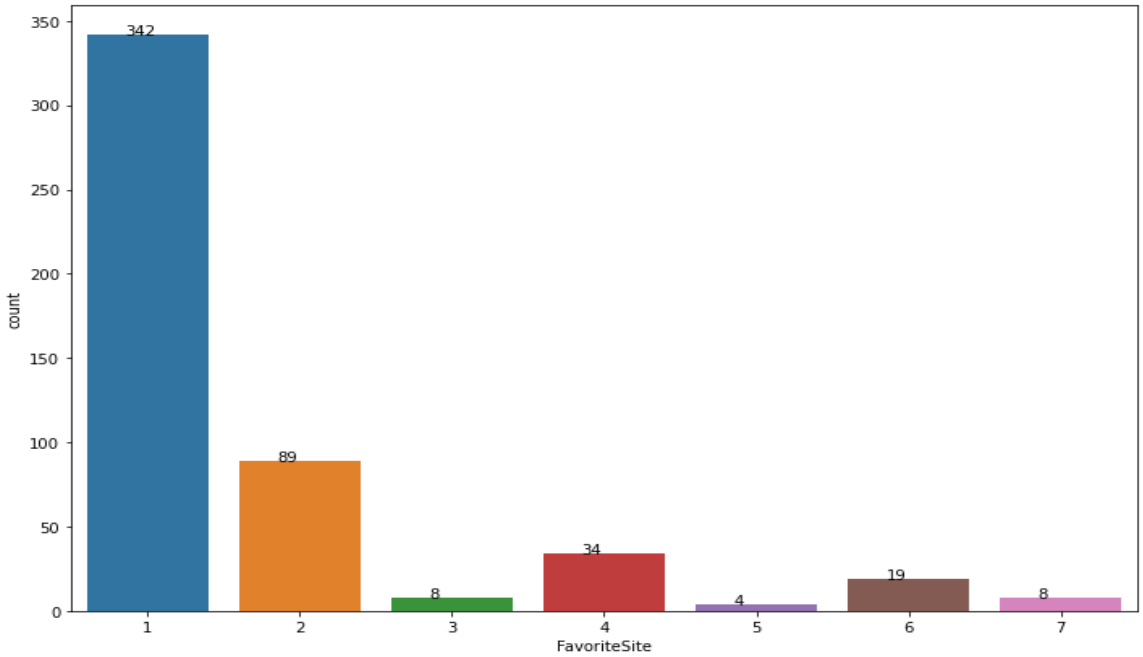


Fig 3.4.8: Favorite site analysis.

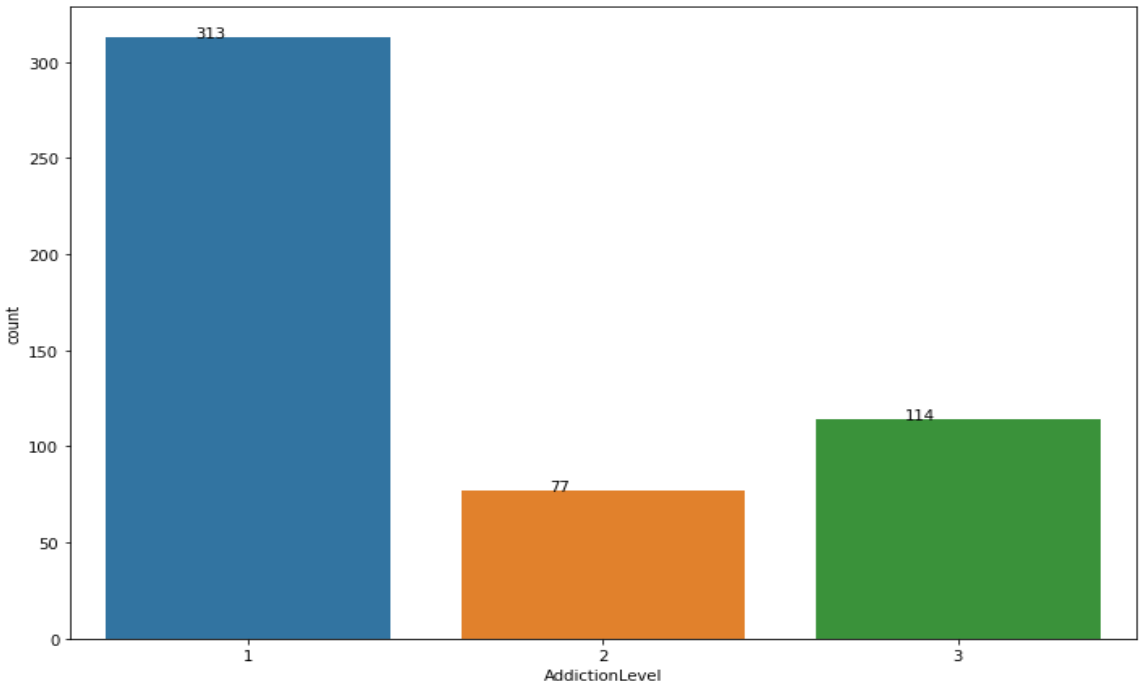


Figure 3.4.9: Histogram of addiction level due to COVID-19

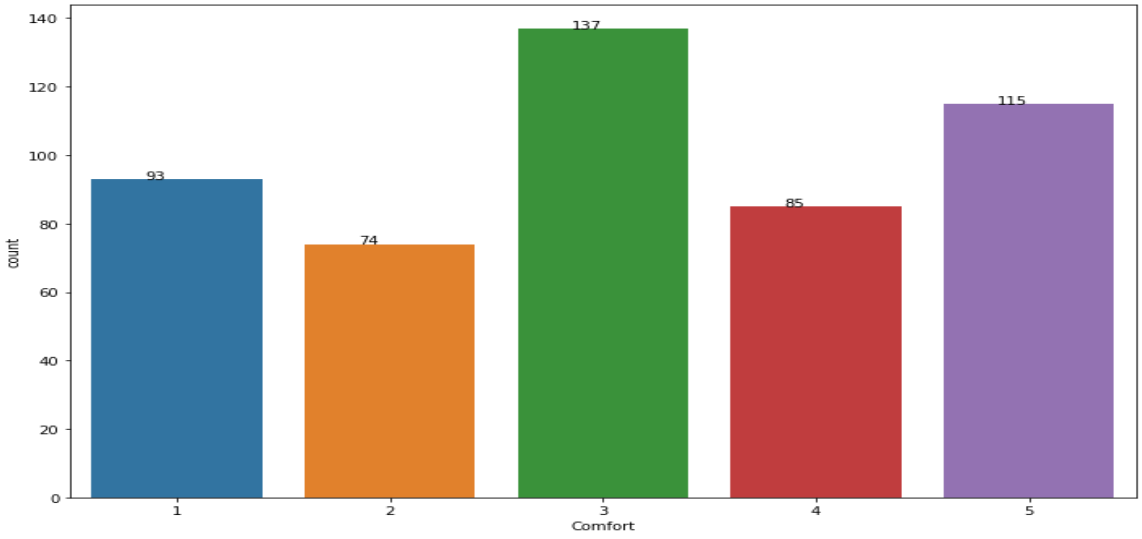


Figure 3.4.10: Histogram of comfort using social media analysis

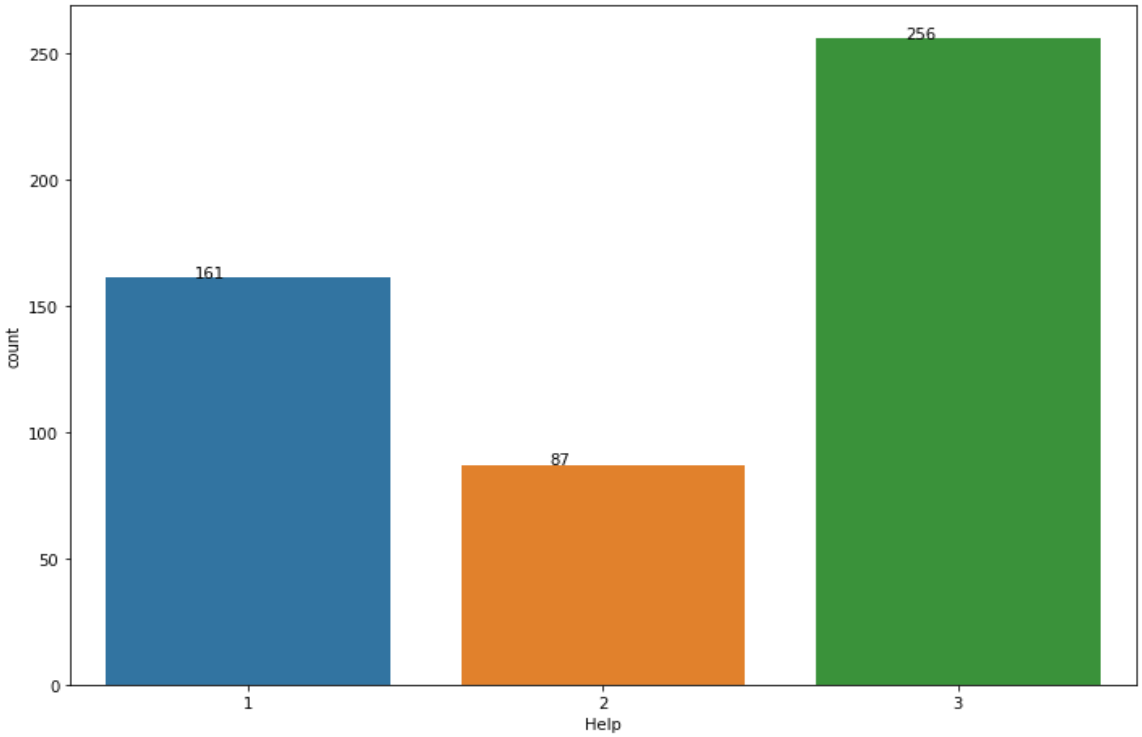


Fig 3.4.11: Histogram of feature help

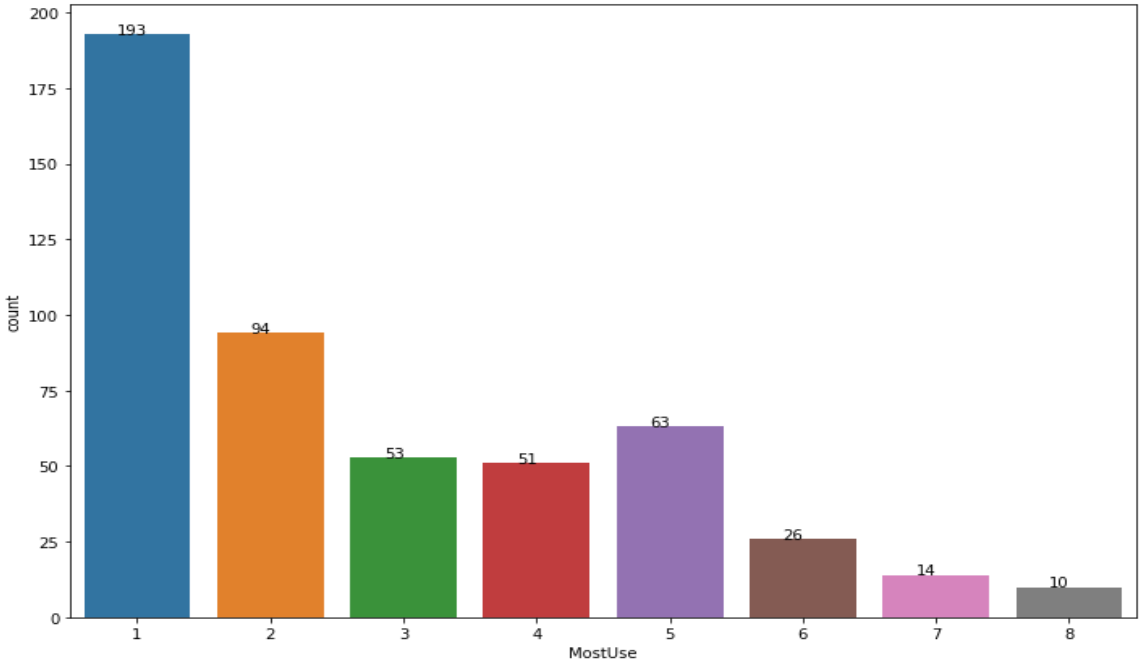


Figure 3.4.12: Histogram of most use time

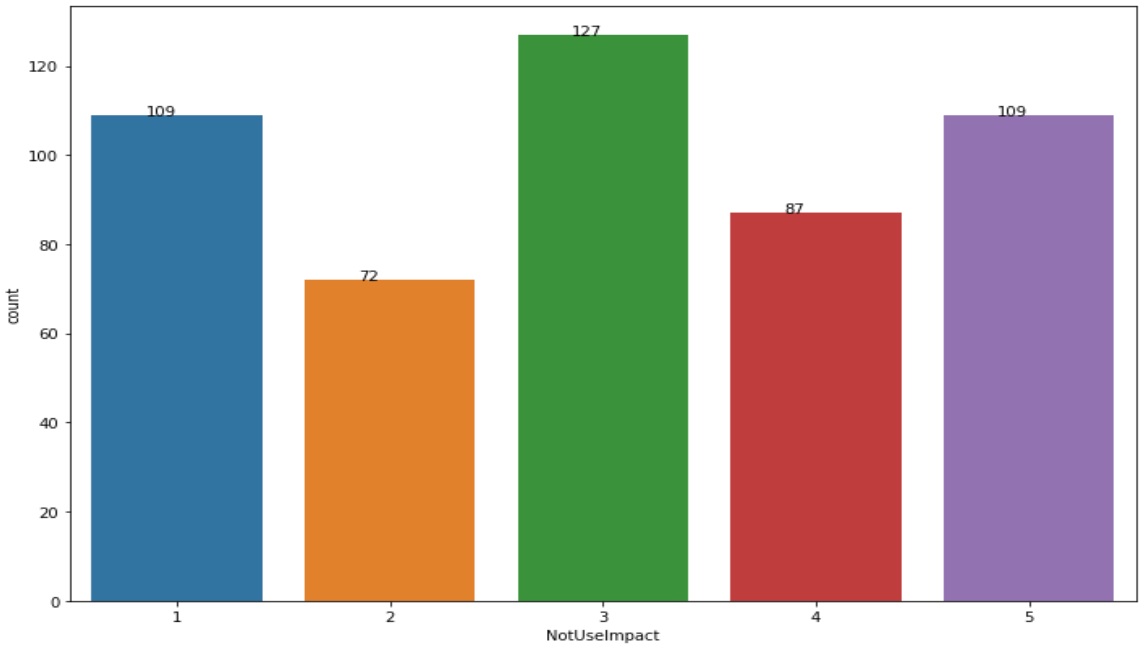


Figure 3.4.13: Histogram of not use impact

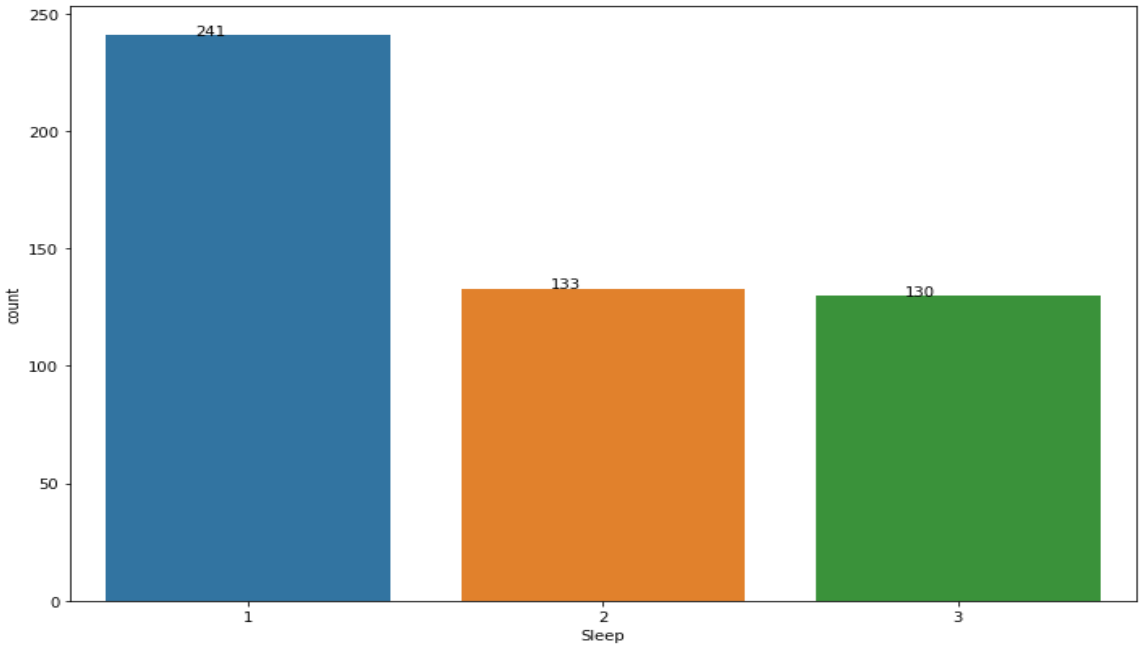


Figure 3.4.14: Histogram of impact on sleep

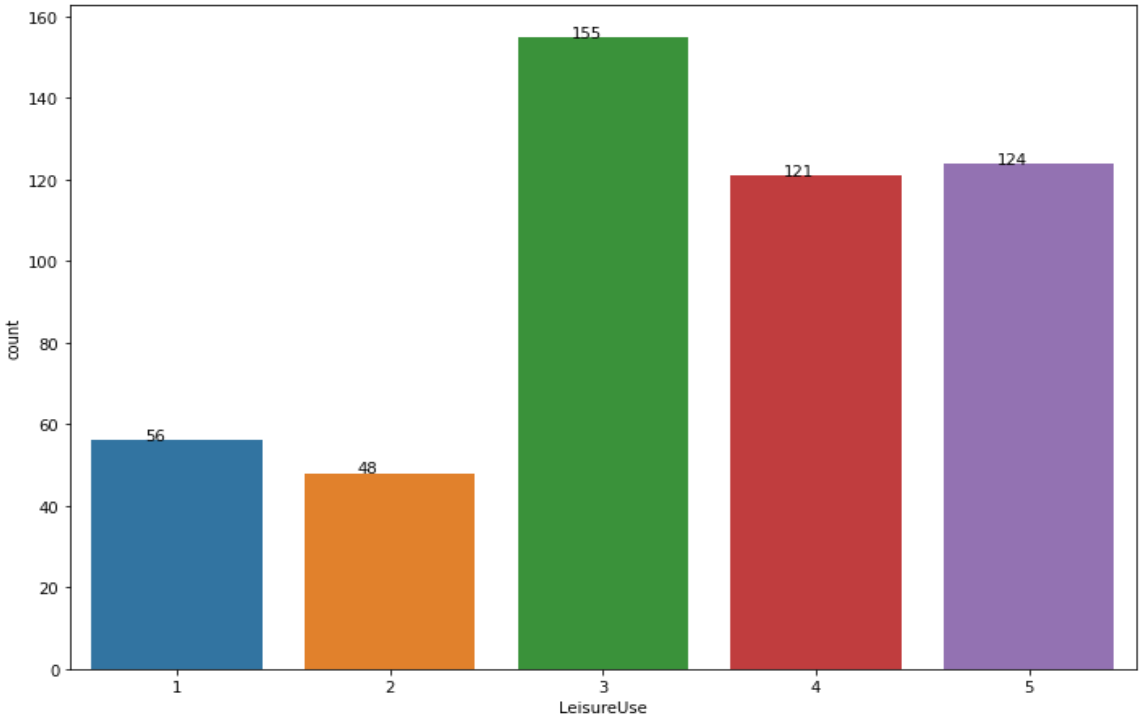


Figure 3.4.15: Histogram of leisure use analysis.

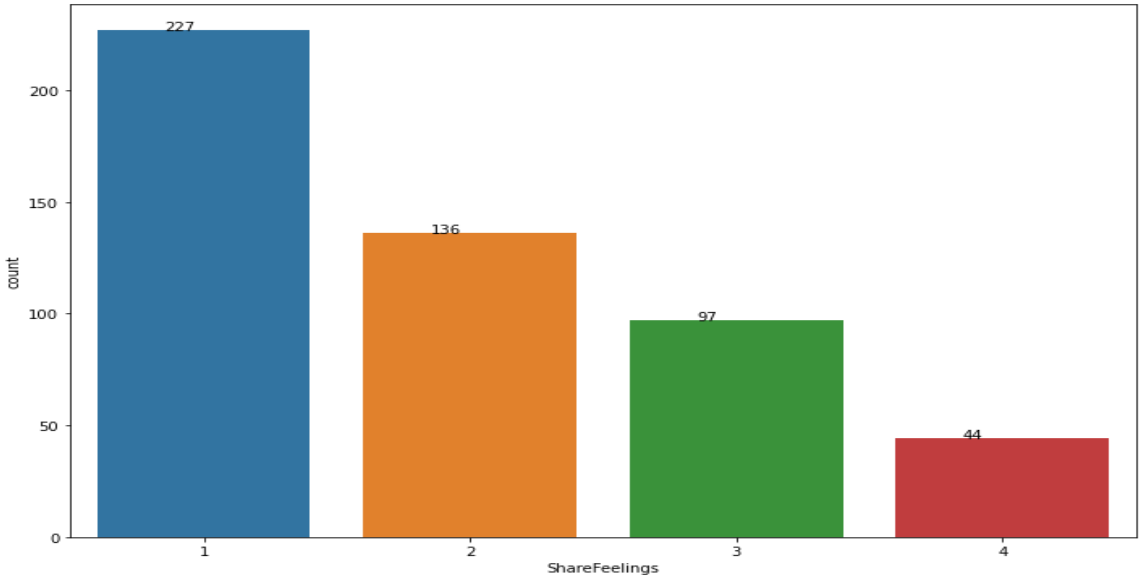


Figure 3.4.16: Histogram of share feelings analysis

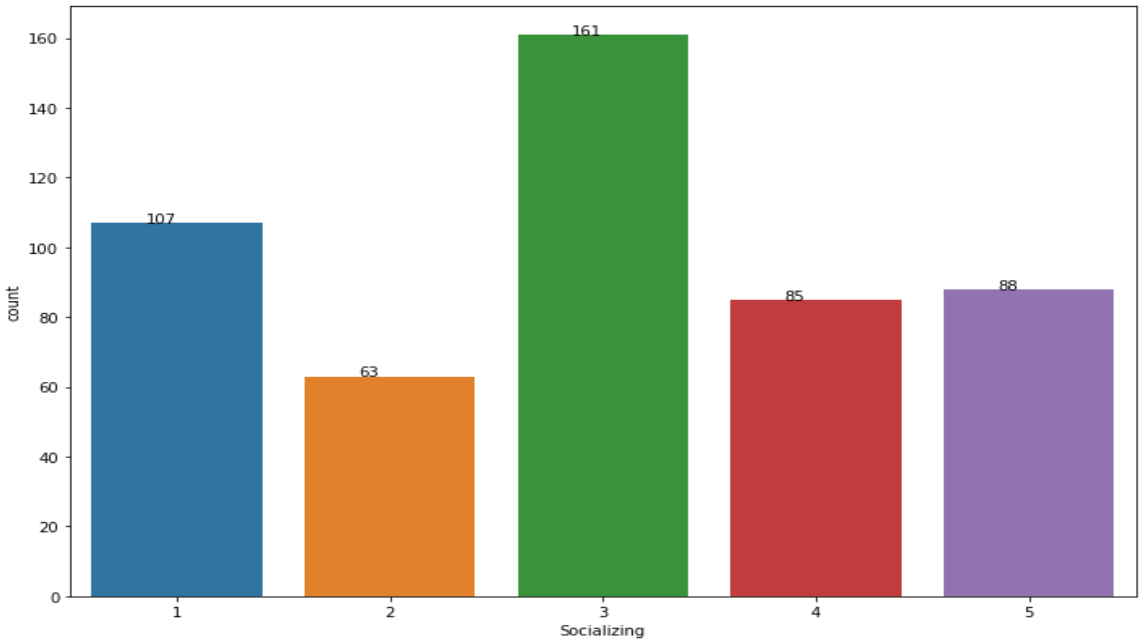


Figure 3.4.17: Histogram of socializing analysis

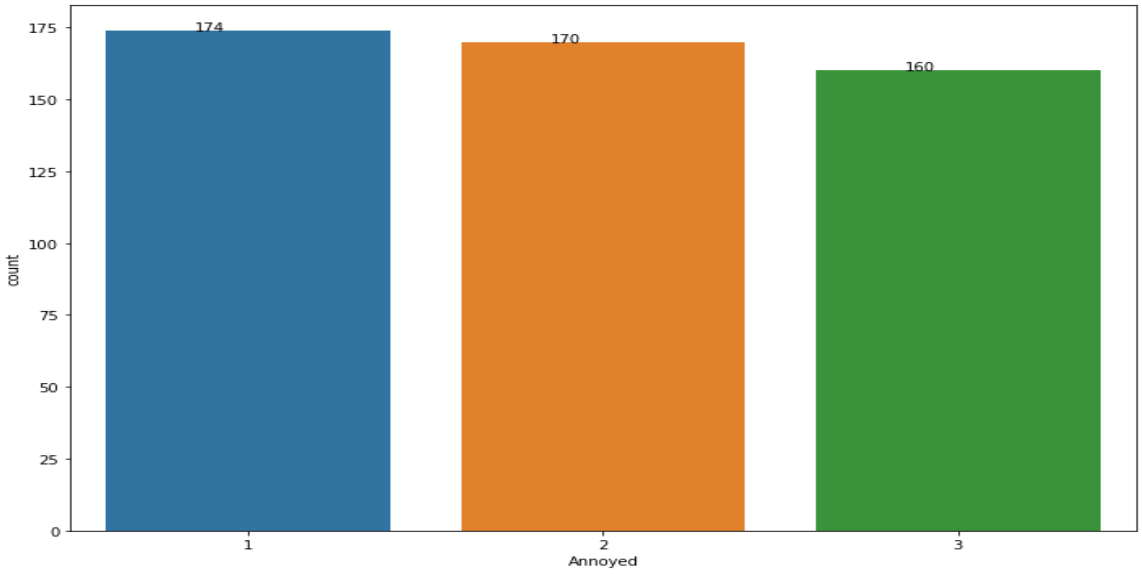


Figure 3.4.18: Histogram of annoyed feature analysis.

3.4.5 Statistical Analysis of Target Feature

In our target feature 353 people were under no addiction category, 81, 19 and 51 people were under general addiction, moderate addiction and severe addiction categories respectively. The statistics is shown in Figure 3.4.19.

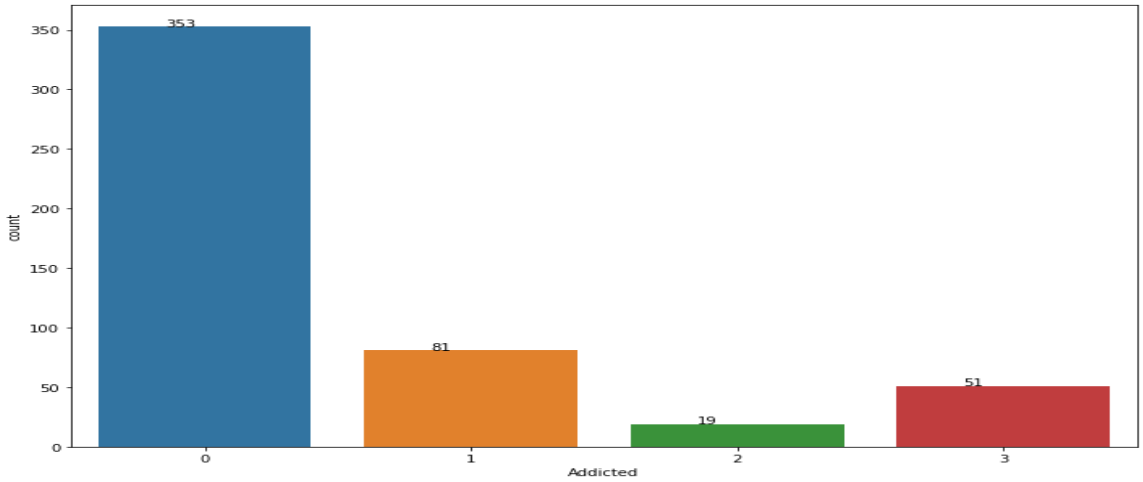


Figure 3.4.19: Histogram of addiction level analysis

3.5 Implementation Requirements

We need data mining tools, data processing tools, data storing tools to implement our work. We collect data through Google forms and using handwritten forms. We created data sets with Microsoft Excel. For data preprocessing and algorithms implementation, we used – "Anaconda Navigator and Jupyter Notebook".

Anaconda Navigator is one kind of graphical user interface for the desktop which allows users to launch application and conda packages, environment and channel without any command-line command and Anaconda has completed and open-source data science packages [28].

3.5.1 Methods and Algorithms Description

Principal Component Analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set [27].

Extremely Randomized Trees Classifier (Extra Trees Classifier) is a type of ensemble learning technique which aggregates the results of multiple de-correlated decision trees collected in a "forest" to output its classification result [29].

Logistic regression is a classification algorithm, used when the value of the target variable is categorical in nature; Logistic regression used logistic function and this Logistic function is called a sigmoid function; an S-shaped curve takes the real values and put them between 0 to 1 [30].

k-nearest neighbors (*k*-NN) is a simple supervised machine-learning algorithm. Classification and regression problems can be explained with the kNN algorithm. *k*-NN algorithm memorizes the training observation for classifying the hidden test data. *k*-NN algorithm grabs similar things that exist in a close neighborhood. [31]

Support vector machine is a supervised machine-learning algorithm. This is also used for both classification and regression problems. Data objects are placed in n-dimensional space and the values of the features are presented the specific coordinate. It generates the most homogeneous points in each subsection that is why it is called hyperplane [31].

Naïve Bayes is one of the oldest algorithms of machine learning. This algorithm is based on Bayes theorem and basic statistics. Class probabilities and conditional probabilities are used in the Naïve Bayes model. It extends attributes using Gaussian distribution [31].

Decision tree is a tree-based model. It distributing the features into the smaller section with similar response value using splitting rules. The divide-and-conquer method uses for making the tree diagram. Decision tree needs a small pre-processing and it can easily control the categorical features without preprocessing [31].

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique; it can be used for both Classification and Regression problems in ML; it is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model [32].

XGBoost classifier is a Machine learning algorithm that is applied for structured and tabular data; XGBoost is an implementation of gradient boosted decision trees designed for speed and performance; XGBoost is an extreme gradient boost algorithm and that means it's a big Machine learning algorithm with lots of parts; XGBoost works with large, complicated datasets; XGBoost is an ensemble modelling technique [33].

Univariate feature selection examines each feature individually to determine the strength of the relationship of the feature with the response variable and these methods are simple to run and understand and are in general particularly good for gaining a better understanding of data [34].

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

This study is conducted for analyzing the social media addiction of people during the Covid-19 pandemic. In this study, information is collected from 504 people of several ages and occupations. After performing different pre-processing techniques on the dataset, seven different methods are used to train seven ML classifiers. Among seven methods, six methods are utilized for selecting features such as principal component analysis (PCA) with 85%, 90%, and 95% variance, correlation, feature importance, and univariate selection. In-depth analysis of different methods and classifiers played a crucial role in this study in finding the best method and classifier for this task.

4.2 Experimental Results & Analysis

In this chapter, the classification performance of seven ML classifiers such as logistic regression (LR), Naive Bayes (NB), XGBoost, support vector machine (SVM), k-nearest neighbors algorithm (k-NN), random forest (RF), and decision tree (DT) is demonstrated, which are evaluated using 20% instances of the dataset. Moreover, we also compared the performance of using Feature Selection Processes (FSP) in this chapter.

4.2.1 FSP I

In FSP I, all features such as Age, Gender, Marital, Occupation, Education, AverageHours, FavoriteSite, AddictionLevel, Comfort, Help, MostUse, NotUseImpact, Sleep, LeisureUse, ShareFeelings, Socializing, and Annoyed were used to train and test seven ML classifiers. In this method, both LR and RF have obtained 94.05% accuracy which was higher than other classes. However, NB achieved very less accuracy than

others, 21.78%. The classification performance of seven ML classifiers is given below in Figure 4.1.

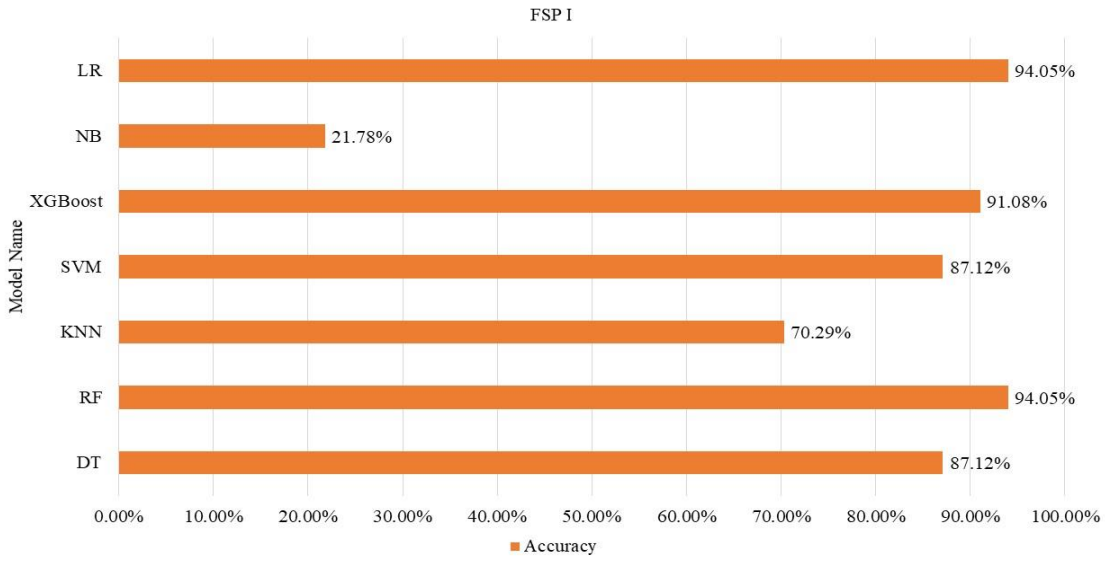


Figure 4.1: The classification performance of seven ML classifiers in the FSP I method.

4.2.2 FSP II

In FSP II, PCA was applied to seventeen features with 85% variance. After that, nine features were selected for training ML classifiers. RF obtained the highest accuracy in this method, it achieved 80.19% accuracy. On the other hand, the performance of the NB classifier was also increased significantly than the previous method, FSP I, and the variance ratio of nine selected features are given below in Figure 4.2.

```
In [7]: pca.explained_variance_ratio_

Out[7]: array([0.26766351, 0.17345397, 0.10718035, 0.06724876, 0.06365893,
               0.05963776, 0.05399766, 0.03847635, 0.03284119])
```

Figure 4.2: The variance ratio of nine selected features using PCA with 85% variance.

The classification performance of seven ML classifiers in FSP II is given below in Figure 4.3.

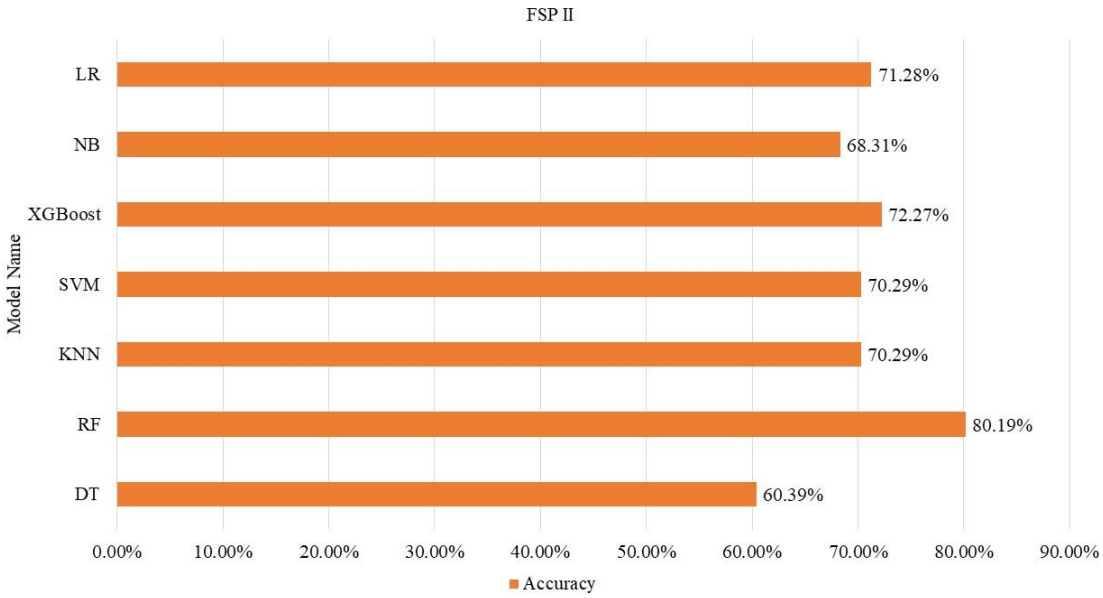


Figure 4.3: The classification performance of seven ML classifiers in the FSP II method.

4.2.3 FSP III

PCA was utilized in FSP III with 90% variance, where 11 features were chosen to train ML classifiers. In this method, LR performed better than other classifiers, obtained 91.08% accuracy. On the other hand, k-NN acquired less accuracy than other classifiers, 70.29%. The variance ratio of nine selected features are given below in Figure 4.4.

```
In [7]: pca.explained_variance_ratio_

Out[7]: array([0.26766351, 0.17345397, 0.10718035, 0.06724876, 0.06365893,
               0.05963776, 0.05399766, 0.03847635, 0.03284119, 0.03148669,
               0.02882276])
```

Figure 4.4: The variance ratio of eleven selected features using PCA with 90% variance.

The classification performance of seven ML classifiers in FSP III is given below in Figure 4.5.

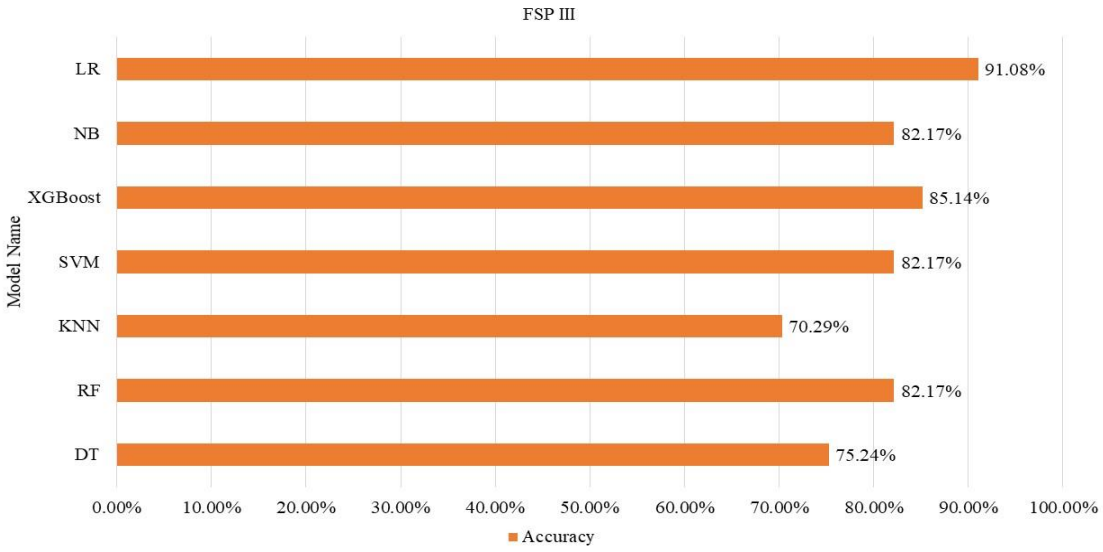


Figure 4.5: The classification performance of seven ML classifiers in the FSP III method.

4.2.4 FSP IV

In FSP IV, 95% variance was used during applying PCA to seventeen features and thirteen features were selected to train ML classifiers. LR performed better than other classifiers in this method, which obtained 94.05% accuracy. After LR, the XGBoost classifier obtained the second highest accuracy in this method and k-NN acquired less accuracy than others. The variance ratio of thirteen selected features are given below in Figure 4.6.

```
In [7]: pca.explained_variance_ratio_
Out[7]: array([0.26766351, 0.17345397, 0.10718035, 0.06724876, 0.06365893,
               0.05963776, 0.05399766, 0.03847635, 0.03284119, 0.03148669,
               0.02882276, 0.02466362, 0.01917503])
```

Figure 4.6: The variance ratio of thirteen selected features using PCA with 95% variance.

The classification performance of seven ML classifiers in FSP IV is given in Figure 4.7.

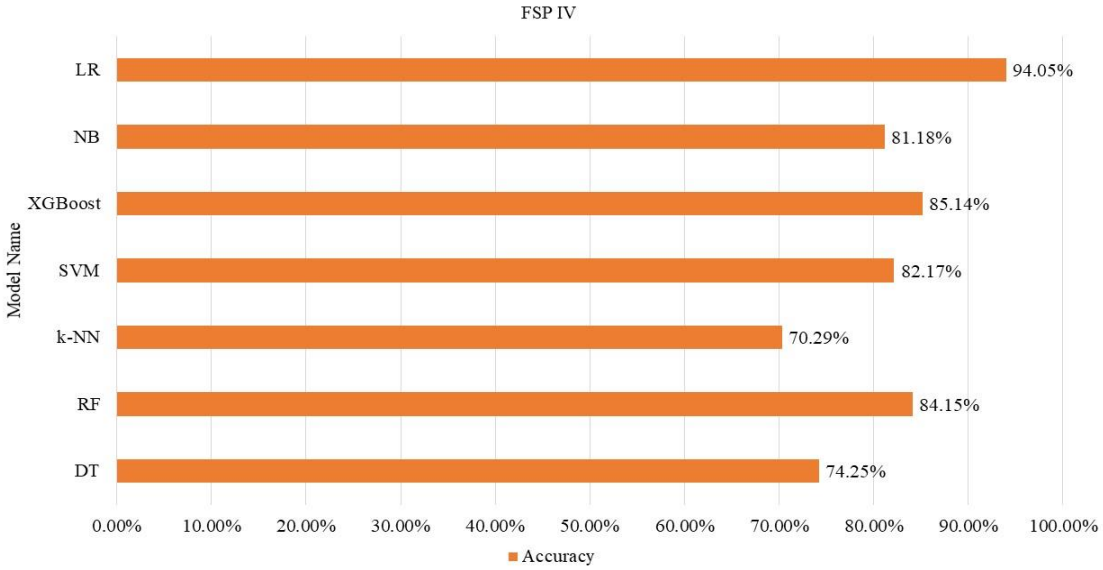


Figure 4.7: The classification performance of seven ML classifiers in the FSP IV method.

4.2.5 FSP V

In FSP V, ten highly important features were selected using the extra tree classifier, which is given below in Figure 4.8.

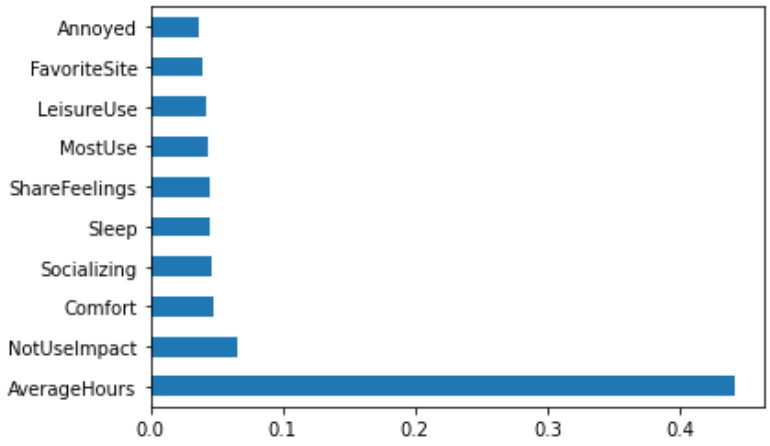


Fig 4.8: Ten important features which were found by the FSP V method.

In this method, XGBoost achieved higher accuracy than other classifiers which obtained 92.07% accuracy. RF also performed well in this method that obtained 90.09% accuracy and k-NN classifier obtained less accuracy than others. The classification performance of seven ML classifiers in FSP V is given below in Figure 4.9.

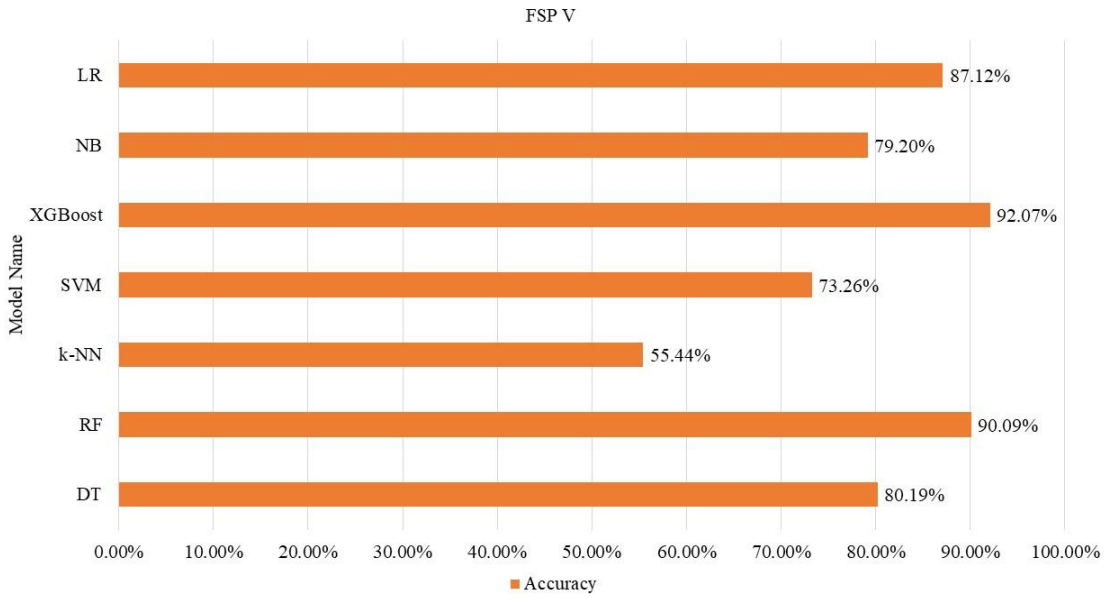


Fig 4.9: The classification performance of seven ML classifiers in the FSP V method.

4.2.6 FSP VI

In FSP VI, four highly pairwise correlated features such as Education, NotUseImpact, Occupation, and Socializing were removed based on pandas dataframe corr(). We also checked the correlation among features using a seaborn library, called as a heatmap. In this method, XGBoost achieved higher accuracy than other classifiers that obtained 91.08% accuracy. On the other hand, the accuracy of k-NN was less than other classifiers, 55.44%. SVM and NB also acquired less accuracy in this method.

The classification performance of seven ML classifiers in FSP V is given below in Figure 4.10.

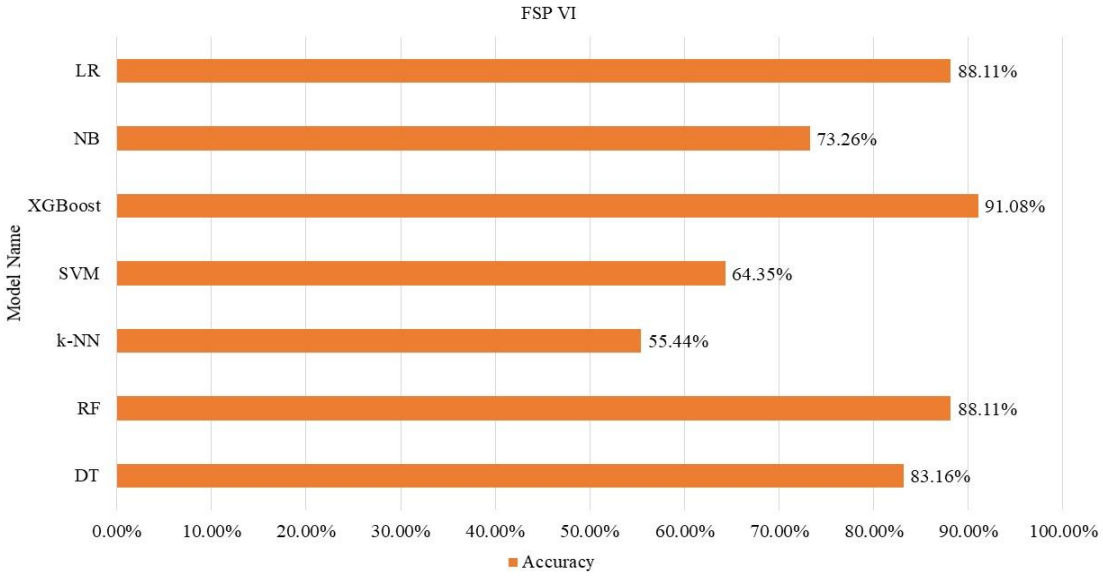


Figure 4.10: The classification performance of seven ML classifiers in the FSP VI method.

4.2.7 FSP VII

In FSP VII, ten most important features were chosen to train the ML classifiers using univariate selection, which uses SelectKBest class. Selected features using the FSP VII method are given in below Figure 4.11 with their importance score.

	Features	Score
5	AverageHours	82.924171
11	NotUseImpact	18.804866
12	Sleep	14.905779
15	Socializing	14.273983
8	Comfort	12.092141
14	ShareFeelings	9.455446
10	MostUse	7.842295
13	LeisureUse	6.864627
16	Annoyed	5.765033
6	FavoriteSite	5.284434

Figure 4.11: Ten important features which were found by the FSP VII method.

In our last method, XGBoost achieved 92.07% accuracy, which was higher than others. After XGBoost, both RF and LR classifiers performed well in this method, and obtained 87.12% accuracy. The classification performance of seven ML classifiers in FSP V is given below in Figure 4.12.

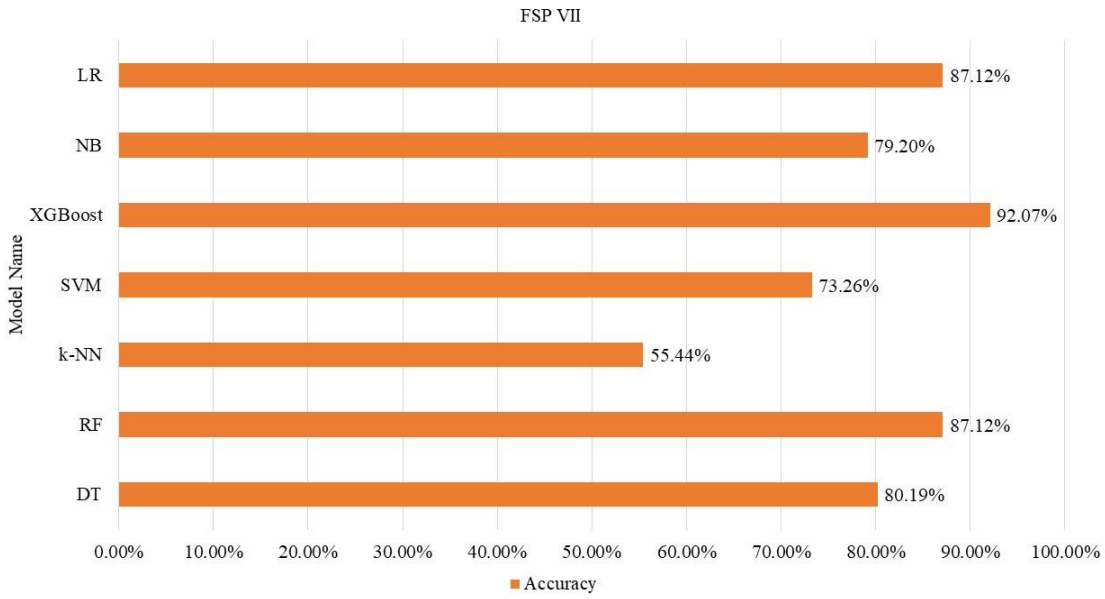


Figure 4.12: The classification performance of seven ML classifiers in the FSP VII method.

4.3 Comparative Analysis

In this study, the performance of seven classifiers was evaluated using different methods. In FSP I, both LR and RF classifiers acquired the highest accuracy but NB acquired the lowest accuracy in this study. The performance of the NB classifier was increased after using PCA, extra tree classifier, and univariate selection method. In FSP III, NB obtained 82.17% accuracy which was higher than other methods. On the other hand, the performance of the k-NN classifier was the same in in FSP V, VI, and VII methods, which was 55.44%. But, the performance of the k-NN classifier was high in FSP I to FSP IV, which was 70.29%. In FSP V and VII, the XGBoost classifier obtained 92.07% accuracy but it achieved 72.27% accuracy in FSP II. In FSP I, DT performed better than

other methods, which obtained 87.12% accuracy, and in FSP II, DT obtained the lowest accuracy than other methods. In FSP I, all features were used, and in FSP II, nine features were chosen for training using PCA with 85% variance. In this study, the FSP IV method performed better than other methods and the lowest and highest accuracy of this method was 94.05% and 70.29%, respectively. On the other hand, LR, and RF classifiers performed better than other classifiers. Moreover, LR obtained the highest accuracy both in FSP I and FSP IV. The performance of seven methods according to their highest accuracy is given below in Figure 4.13.

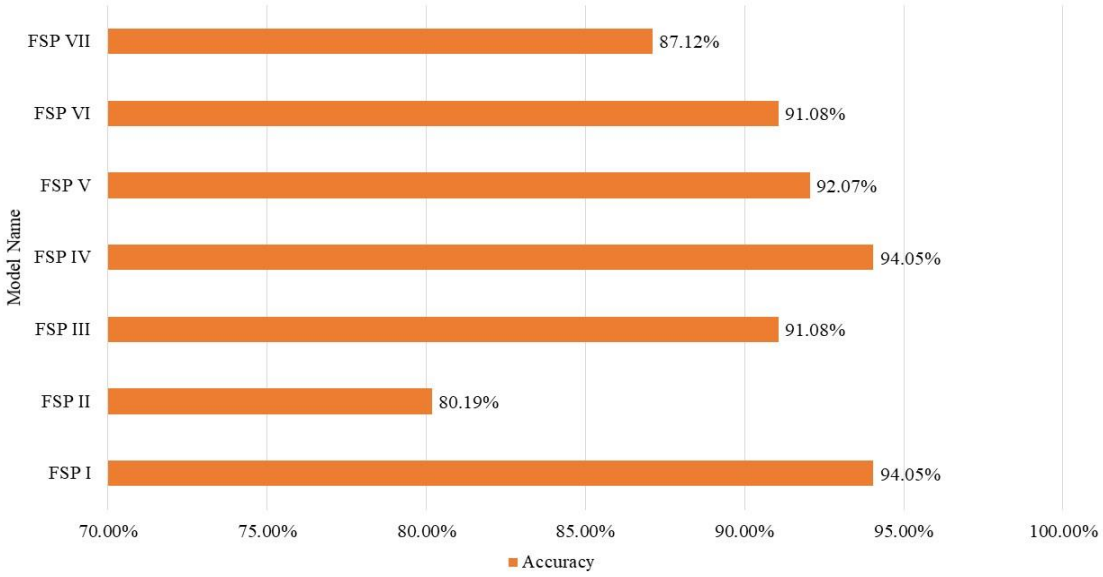


Figure 4.13: The classification performance of seven methods.

In this study, LR was selected as the best classifier as it obtained the highest accuracy in both FSP I and FSP IV methods. The performance of the best classifier of this study was also compared with other studies which were conducted in the literature and are given below in Table 4.1.

TABLE 4.1: COMPARISON OF THE PROPOSED CLASSIFIER WITH OTHER STUDIES.

Method/Work Done	Object(s) Deal with	Problem Domain	Sample size	Size of Feature set	Algorithm	Accuracy
This work	Social media addiction	Prediction	504	18	LR	94.05%
Sabbir Alam et al.	Emotion Analysis of Bangladeshi People during Covid-19 in Social Media	Prediction	1120	NM	CNN	97.24%
Wahyu Rahardjo et al.	Instagram addiction	Perception	434	NM	Snowball sampling and regression techniques	56.9% (affected)
Johura Khatun et al.	Betel nut addiction	Detection	1001	19	RF	99.00%
Zahirul et al.	Facebook addiction	Prediction	1001	24	SVM	85.00%
CIPLAK and Ersun	Social Media Addiction: Time spent, Happiness,	Prediction	239	NM	Frequency, percentage, independent sample t-test,	Time spent, Narcissism: Positive

	Narcissism				Pearson correlation coefficient, and stepwise regression	Happiness: Negative
Md. Rashedul Hasan	Social Media Addiction on Young People during COVID 19	Analysis	520	NM	SPSS v26	NM
Aydin et al.	Social Media Addiction and Depression among university students of Turkey	Analysis	419	NM	Typical Screening Model	NM
İmran and Mehmet	Social Media Addiction	Analysis	665	NM	Descriptive Statistics, Kruskal Wallis Test, One-Way ANOVA test, Factor Analysis, and Correlation	NM
Haydar Hoşgöret et al.	Relationship between addiction to	Analysis	205	NM	IBM SPSS V.22.0, t test,	NM

	social media and job among nurses				ANOVA, Pearson's correlation, and simple regression analysis	
Rahmatullah HAAND and Ahmad Zia ELHAM	Addiction of social media of university students in Afghanistan	Analysis	384	NM	Probability sampling method	NM
Mustafa USLU	Addiction of social media usage among Turkish people	Analysis	5176	NM	Descriptive study in survey model	NM

NM[1]: Not Mentioned.

4.4 Discussion

Some studies were conducted for analyzing social media addiction under several conditions in the literature. But, no study was conducted for analyzing the social media addiction of people during the Covid-19 pandemic**.** For analyzing social media addiction of people during the Covid-19 pandemic, this study was conducted using the information of 504 people. We have used seven different methods and classifiers in this study. According to the result of experimental studies, PCA with 95% variance performed better than other methods. Moreover, the LR classifier showed significant performance in the two methods. The performance of some classifiers was very close to each other due to fewer instances in the dataset, it can be overcome by increasing the number of instances. However, in this study, LR showed a remarkable performance that obtained 94.05%

accuracy, which strongly validates the efficiency of used method and classifier. In-depth analysis of different method performance also demonstrated clearly efficiency of methods in this study.

CHAPTER 5

Impact on Society, Environment, and Sustainability

5.1 Impact on Society

The machine learning model to predict social media addiction will have a good impact on society. People of various religions and classes must work together to live in a society since humans are social beings. A social media addict becomes an impediment in their journey together. For a variety of reasons, people might get addicted to social networking. In the preceding discussion, we saw that individuals are addicted to social media for entertainment and socializing, and that many people use social media to avoid various traumas such as despair, irritation, and stress. This is how our society's younger generation is gradually becoming addicted to social media. Parents and guardians should be aware of and concerned about their children at all times. It is the obligation of the parents to provide their children time, to use them in a friendly manner, and to monitor their actions. If a parent is ever in question, he or she can utilize this model to offer the knowledge and statistics needed to determine whether or not their child is addicted to social media. We can protect our young society in this way, preventing them from being addicted to social media. We believe that our social media addiction prediction model will be beneficial to society as a whole.

5.2 Impact on Environment

Our model is not detrimental to the environment in any way. This model does not require any chemicals, combustibles, or organic acids to work. As a result, this model will have no negative consequences for the ecosystem or biodiversity. By influencing others to use social media, a social media junkie might have an impact on others. As a result, it's critical to identify addicts and take the required actions to help them recover. We will be able to live our lives in peace as a result of this, and the number of crimes related to

social media usage will diminish. As a result, our concept is advantageous to individuals who are addicted to social media and want to escape the clutches.

5.3 Ethical Aspects

This addiction prediction model is not unethical or infringes on human rights in any manner. There will be no privacy issues because the model does not collect any personal information, such as name, identification, or location. This model does not infringe on a person's right to enjoy or use, but it does play a part in raising awareness. The risk of social media addiction prediction model was designed with all types of restrictions in mind, as well as privacy and confidentiality concerns. As a result, the model for predicting social media addiction can be easily maintained utilizing machine-learning technologies.

5.4 Sustainability Plan

The community, financial, and organizational aspects of the sustainability strategy are all important. The Sustainability Plan gives us a realistic picture of how a project will run and what the project's future plans are. The goal of our model cohort is to identify those who have a proclivity for social media addiction. This model must be tailored to make it simple for people to adjust, and it is critical to remember that using this model does not imply that people are inferior. This model can be used by psychiatrists, psychologists, and mental health organizations to speed up their work.

CHAPTER 6

Summary, Conclusion, Recommendation, and Implementation for Future Research

6.1 Summary of the Study

The addiction of social media is harmful for the society. The addicted persons are not mentally stable. They suffer from various mental problems such as frustration, depression, irritation etc. Young people mostly suffer from social media addiction, they could not concentrate on their studies and works. People can't use their valuable time properly because of social media. Due to the pandemic of covid-19 the addiction of social media has increased rapidly. We have used machine learning approach for predicting the addiction of social media during to covid-19.

6.2 Limitation and Conclusions

We have used machine learning approach for detecting the addiction of social media due to covid-19. We have some limits in our work and model. We face problems when we collect data due to lockdown. We can't train our model with huge data set, the larger the data set the prediction will be more accurate.

6.3 Implication for Further Study

In near future we'll try to expand our dataset and we'll do better result and more accuracy, and make a software or web application so that anyone can measure the level of addiction. Besides we will do research on individual social sites as well as different kind of addiction.

APPENDIX

Abbreviation

FSP = Feature Selection Process
PCA = Principal Component Analysis
LR = Logistic Regression
k-NN = *k*-nearest neighbors.
SVM = Support Vector Machine.
NB = Naïve Bayes.
DT = Decision Tree.
RF = Random Forest.
XGBoost = eXtreme Gradient Boosting.

Appendix: Research Reflection

We knew very little about machine learning and artificial intelligence detection and prediction when we started this research project. Our supervisor was a kind person who was always willing to help. He provided us with invaluable advice and was quite helpful. We learnt numerous new techniques, new knowledge, how to employ algorithms, and how to work with various methodologies during the course of our research. The Anaconda-navigator and Jupyter notebook, as well as the Python programming language, were also new to me. There were some difficulties at first, but we progressively became more comfortable with Anaconda-navigator, Jupyter notebook, and Python.
Finally, conducting the research gave us courage and motivated us to do more in the future.

REFERENCES

[1] Article from the book Change: 19 Key Essays on How the Internet Is Changing Our Lives, "How the Internet Has Changed Everyday Life" by Zaryn Dentzel, Tuenti, Madrid, Spain, available at <https://www.bbvaopenmind.com/en/articles/internet-changed-everyday-life/>, accessed on October 2021

[2] "Social Media" by Maya Dollarhide, updated August 31, 2021, reviewed by Amy Drury, Investopedia. Available: https://www.investopedia.com/terms/s/social-media.asp

[3] "Global Social Media Stats", Datareportal, available at <https://datareportal.com/social-media-users>, last accessed on November 2021

[4] Md. Tanvir Hossain et al, "Social and electronic media exposure and generalized anxiety disorder among people during COVID-19 outbreak in Bangladesh: A preliminary observation", published: September 11, 2020, doi: https://doi.org/10.1371/journal.pone.0238974.

[5] "Social Media Addiction", AddictionCentre, available at https://www.addictioncenter.com/drugs/social-media-addiction/, last accessed on September 2021

[6] "Social Media and Mental Health", HelpGuide, available at <<https://www.helpguide.org/articles/mental-health/social-media-and-mental-health.htm>>, last accessed on November 14, 2021.

[7] Dhaka Tribune, "Cyberbullying: How to report crimes online?" by Arifur Rahman Rabbi, published at 11:56 am August 24th, 2020, available at <<https://www.dhakatribune.com/bangladesh/2020/08/24/cyberbullying-how-to-report-crimes-online>>

[8] "Human Traffickers Use Social Media to Target Children" by Dr. Jarrod Sadulski, published on April 17, 2020. Available: https://amuedge.com/human-traffickers-use-social-media-to-target-children/

[9]WEBROOT, "Internet Addiction: How Much is Too Much Time Online?" Available: https://www.webroot.com/us/en/resources/tips-articles/internet-addiction-too-much-time-online

[10]Sleep Foundation, "Sleep and Social Media" by Rob Newsom, updated November 13, 2020. Available: https://www.sleepfoundation.org/how-sleep-works/sleep-and-social-media

[11] "Social Media and Mental Health" by Lawrence Robinson and Melinda Smith, M.A., last updated: October 2021. Available: https://www.helpguide.org/articles/mental-health/social-media-and-mental-health.htm

[12] "7 Signs of Social Media Addiction", posted at 22:52h in Take Control, Tech Addiction by Isabella Ward. Available: https://www.itstimetologoff.com/2020/07/14/7-signs-of-social-media-addiction/

[13] Future of Work, “12 Most Obvious Signs you are Addicted to Social Media” by Daniel Newman, July 26, 2011. Available: https://fowmedia.com/12-most-obvious-signs-you-are-addicted-to-social-media/

[14] Md. Sabbir Alam Pran, Md. Rafiuzzaman Bhuiyan, Syed Akhter Hossain, and Sheikh Abujar, “Analysis of Bangladeshi People’s Emotion during Covid-19 in Social Media Using Deep Learning”, 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, DOI**:** 10.1109/ICCCNT49239.2020.9225500

[15] Wahyu Rahardjo, and Indah Mulyani, “Instagram addiction in teenagers: The role of type D personality, self-esteem, and fear of missing out”, May 2020 Psikohumaniora Jurnal Penelitian Psikologi 5(1):29-44, DOI:10.21580/pjpp.v5i1.4916

[16] Johura Khatun et al, “Betel Nut Addiction Detection Using Machine Learning”, 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, DOI: 10.1109/ICECA49313.2020.9297437

[17] Islam, Md. Zahirul; Jannat and Ziniatul, “Detection of Facebook Addiction Using Machine Learning”. Available: http://dspace.daffodilvarsity.edu.bd:8080/handle/123456789/5136, July 2020

[18] Ersun Çıplak, “Variables Predicting Social Media Addiction: Narcissism, Happiness and Average Weekly Time Spent on Social Media”, published in International Online Journal of Educational Sciences (IOJES), 2020-Volume 12, Issue 5

[19] Rashedul Hasan, “A Study on Social Media Addiction of Young People during COVID-19 in Bangladesh”, Journal: Academia Letters, August 2021.

[20] Serdar Aydin et al, “Investigation of the Effect of Social Media Addiction on Adults with Depression”, Healthcare (Basel). 2021 Apr; 9(4): 450, Published online 2021 Apr 11. doi: 10.3390/healthcare9040450

[21] Imran Aslan and Mehmet Emin Yaşar, “Measuring Social Media Addiction among University Students”, December 2020, DOI:10.5281/zenodo.4429749

[22] Haydar Hoşgör, “Social media addiction and work engagement among nurses”, published: 29 March 2021. Available: https://doi.org/10.1111/ppc.12774

[23] Rahmatullah HAAND, “Social Media Addiction in the Least Developed Countries: a quantitative study among university students in Afghanistan”, published in International Journal of Sport Culture and Science, Year 2021, Volume 9, Issue 2, 233 - 245, 21.07.2021

[24] Mustafa USLU, “The Analysis of Social Media Addiction and Usage in Turkey”, e-ISSN: 2602-2923, Year: 2021, Volume: 6, Issue: 2. Available: https://dergipark.org.tr/en/download/article-file/1753869

[25] Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. Business Horizons, 53(1), 59-68.

[26] “A Step-by-Step Explanation of Principal Component Analysis”, built in Beta. Available: https://builtin.com/data-science/step-step-explanation-principal-component-analysis

[27] “Anaconda Navigator.”[Online], available: <<https://docs.anaconda.com/anaconda/navigator/index.html>>, last accessed on September 9, 2021.

[28] GeeksforGeeks,”ML | Extra Tree Classifier for Feature Selection”. [Online]. Available: https://www.geeksforgeeks.org/ml-extra-tree-classifier-for-feature-selection/

[29] ML Glossary, “Logistic Regression”, [Online], available: https://ml-cheatsheet.readthedocs.io/en/latest/logistic_regression.html

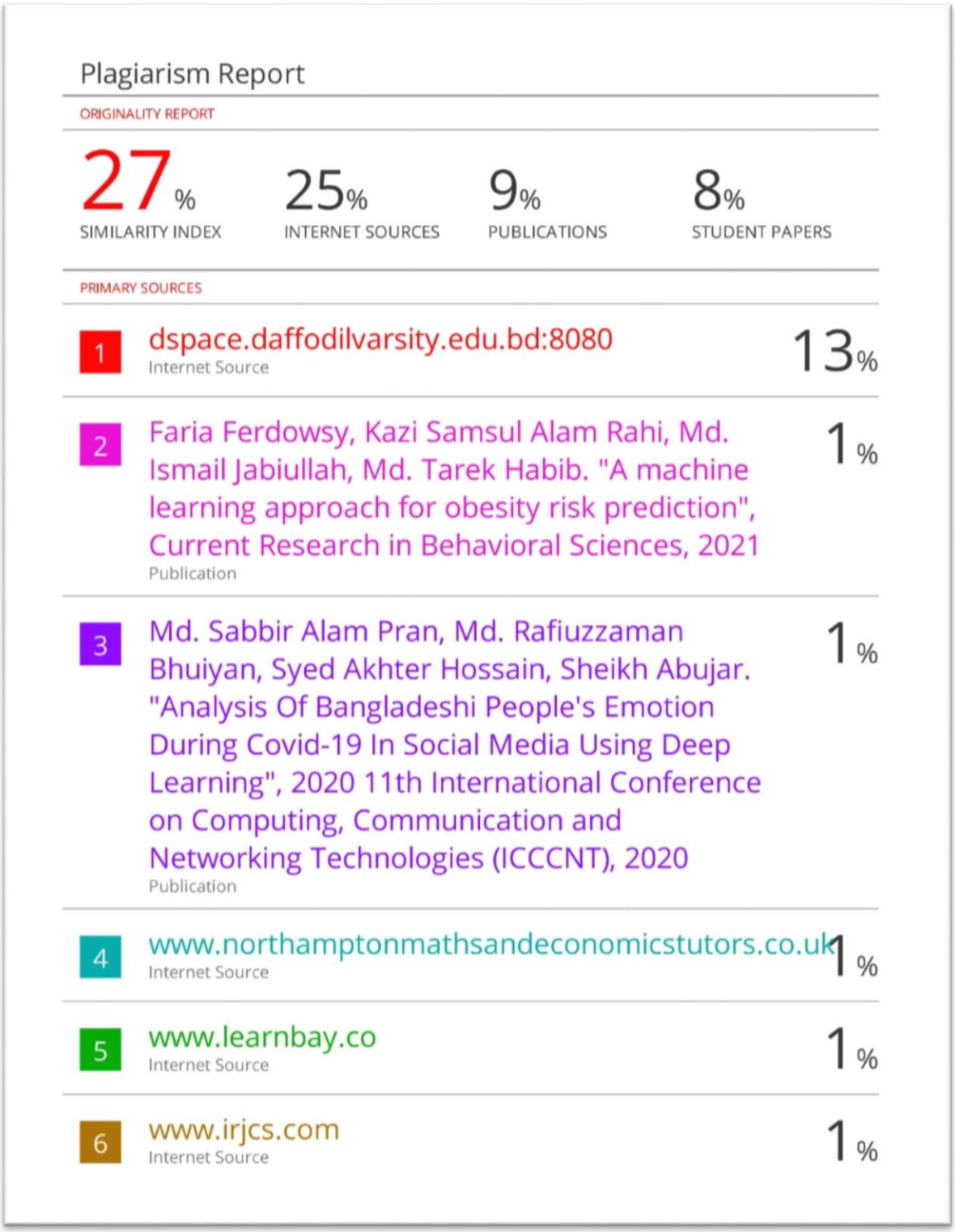
[30] “Machine Learning Classifiers – The Algorithms & How They Work” [Online], available: https://monkeylearn.com/blog/what-is-a-classifier/

[31] “Understanding Random Forest”, available: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

[32] “A Gentle Introduction to XGBoost for Applied Machine Learning”, by Jason Brownlee on August 17, 2016, last updated on February 17, 2021. Available: https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/

[33] “Feature selection – Part I: univariate selection” [Online], available: https://blog.datadive.net/selecting-good-features-part-i-univariate-selection/

[34] Javapoint, “Data Preprocessing in Machine Learning”, accessed on October, 2021. Available: https://www.javatpoint.com/data-preprocessing-machine-learning



Plagiarism Report

ORIGINALITY REPORT

27%	25%	9%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	13%
2	Faria Ferdowsy, Kazi Samsul Alam Rahi, Md. Ismail Jabiullah, Md. Tarek Habib. "A machine learning approach for obesity risk prediction", Current Research in Behavioral Sciences, 2021 Publication	1%
3	Md. Sabbir Alam Pran, Md. Rafiuzzaman Bhuiyan, Syed Akhter Hossain, Sheikh Abujar. "Analysis Of Bangladeshi People's Emotion During Covid-19 In Social Media Using Deep Learning", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020 Publication	1%
4	www.northamptonmathsandeconomicstutors.co.uk Internet Source	1%
5	www.learnbay.co Internet Source	1%
6	www.irjcs.com Internet Source	1%