Analyzing and Predicting the Social Media Addiction Using Machine Learning

By
Hemayet Hossain Tuhin
211-15-4042

Md Abdul Kayum
211-15-4043

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised by

Md. Firoz Hasan
Lecturer (Senior Scale)
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by

Abdullah Al-Amin
Lecturer
Department of Computer Science and EngineeringDaffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

January 12, 2025

APPROVAL

This Project titled **"Analyzing and Predicting the Social Media Addiction Using Machine Learning"**, submitted by Hemayet Hossain Tuhin, ID No: **211-15-4042** and Md Abdul Kayum, ID No: **211-15-4043** 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 12 January, 2025.

BOARD OF EXAMINERS

Dr. Md. Taimur Ahad — **Chairman**
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mushfiqur Rahman — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Rahmatul Kabir Rasel Sarker — **Internal Examiner**
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Sadat Hasan — **External Examiner**
Data Scientist (Senior Principal Officer)
Risk Management Division
BRAC Bank

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md. Firoz Hasan, Lecturer (Senior Scale)**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

11.01.25

Md. Firoz Hasan
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Abdullah Al-Amin
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Hemayet, 11.01.2025

Hemayet Hossain Tuhin
Student ID: 211-15-4042
Department of Computer Science and Engineering
Daffodil International University

11.01.2025

Md Abdul Kayum
Student ID: 211-15-4043
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Md. Firoz Hasan, Lecturer (Senior Scale)**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Machine Learning** carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

The new reality of our modern life is social media. Modern science has given people a better civilization, increased the standard of living, but has taken away all the emotions of life. Social media has become an essential part of our modern lives. It significantly affects how people communicate, work, and interact with each other. Social media offers many benefits, such as increased connectivity and information sharing, but excessive use has led to growing concerns about addiction, which can negatively impact on mental health, productivity, and relationships. Our goal is to predict social media addiction and evaluate its impact on professional and personal life using machine learning algorithm. We used six machine learning classifiers like Random Forest, Extra Tree Classifier, Linear Discriminant Analysis (LDA), Gaussian Naive Bayes, and ensemble approaches including Stacking and AdaBoost to analyze a dataset of 522 samples that collected via structured questionnaires. Which included data from students and employed people. Outperforming the other models, Stacking Ensemble model achieved the highest accuracy of 91.71% in classifying social media addiction behaviors. This study not only provides valuable insights into identifying addiction behaviors, but it also proposes a practical solution in the form of a mobile application based on a superior predictive model. The goal of this application is to help people increase their consciousness and encourage people to adopt good social media practices.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter discusses in detail the importance of social media, its positive and negative effects, and the objectives and methodology of our research. It analyzes the addiction to social media use and its effects and provides practical solutions for raising awareness through a user-friendly mobile application. The chapter emphasizes the main motivation, objectives, and application of the research findings in real life.

1.1 Introduction

Social media has become a vital part of modern life. It has a significant impact on the way people live their lives [1]. People can't spend a single day without using it. In Bangladesh the total number of social media users was 44.70 million in January 2023 [2]. In January 2024, the total number of social media users in Bangladesh was 52.90 million which represents an increase of 18.37% [3]. The increase in frequency use of social media platforms has risen from a mere 36 million users in 2020 showing a 47.22% growth rate [4]. These results depict that, there was progressive increase in the frequency use of the social media platforms.

Social media has provided people with new ways to connect with other people on the internet. People use social media, like online groups, to make their own material and connect with other users [5]. These platforms are made to help people connect with each other and share thoughts, pictures, videos, and other material in a digital setting. It has become an important part of modern communication and has changed how people connect with each other. Businesses use social apps to sell and promote their goods and keep track of what customers are saying about them. The goal of social networking sites, which are online platforms, is to create and sustain relationships between users. Social networking sites come in a variety of forms, each with unique features and objectives. However, in this study, we highlight a few social media platforms that are mostly user-friendly, including Facebook, YouTube, Instagram, TikTok, Imo, Telegram, WhatsApp, Twitter, and Linkin.

1.2 Motivation

Social media has become an integral part of an individual's daily life and the whole digitally developed era. It is essentially a tool for communication and exchange of information and for gaining knowledge. But its uncontrolled and excessive utilization leads to detrimental impacts on individual, society and work domain which includes, but not limited to stress, decreased work efficiency and addiction. Although relevant studies have discussed various aspects of the impact of social media, most of them have been limited to certain platforms or only positive or negative aspects. In addition, a lack of practical solutions has been observed for analyzing user behavior. The main objective of this study is to analyze both positive and negative effects of social media and to measure and increase awareness of users' addiction through a user-friendly mobile application. Through this study, we are not only enriching theoretical knowledge, but also providing a practical solution for users, which will be helpful in improving their social media usage habits. To this end, the inspiration for our research comes from the desire to reduce the negative effects of excessive use of social media and increase users' awareness.

1.3 Objectives

Our objective is to develop a machine learning model for analyzing and predicting the social media addiction with the deployment of model into mobile application. Identify whether social media affects his/her professional or personal life. Comprehensively examine both the positive and negative effects of social media usage on daily life, incorporating insights into user awareness and responses. And combine real statistical data with quantitative and qualitative findings.

1.4 Methodology

In order to predict and analysis social media addiction, we used a machine learning-based methodology in this study. Initially, 522 sample datasets were collected, including information from employee and students. Data was collected using Google Forms, where questions were constructed using 14 features. Less significant characteristics were eliminated from the dataset during the data preparation stage, and categorical data was transformed into numeric-based data using label encoding. After processing, the dataset was ready for various model training. The study used five machine learning models: Gaussian Naïve Bayes, Random Forest, Linear Discriminant Analysis (LDA), Extra Trees, and two ensemble techniques called Stacking and Adabost. Each model's performance was assessed using appropriate measures, including F1-score, recall, accuracy, and precision. The stacking ensemble model had the highest accuracy of 91.71% and outperformed the other models. This model was integrated into an Android application that allows users to enter their

data and receive addiction forecasts. The program uses a Flask API to process data and make predictions. This method has resulted in a successful blend of research and real-world applications, as well as a solid foundation for recognizing social media addiction.

1.5 Project Outcome

Our study will try to develop a machine learning model for the prediction of social media addiction and determine how far it may affect the professional and personal lives of people. We will analyze comprehensively the positive and negative impacts of social media upon daily life, and we will also incorporate real statistical data with quantitative and qualitative findings. Also deploy the machine learning model into a mobile application. The approach will help us understand how social media affects our lives with greater clarity.

1.6 Organization of the Report

This report outlines the development and deployment of a machine learning-based system for analyzing and predicting social media addiction, aiming to provide practical solutions and raise awareness about healthy social media usage habits. The report is organized into six chapters. The first chapter introduces the study, highlighting its motivation, objectives, and the methodology employed. The second chapter provides a comprehensive background, including a literature review and a gap analysis, to establish the need for this research. In the third chapter, we detail the methodology and the deployment process, covering data collection, preprocessing, model development, and the integration of the machine learning model into an Android application for real-world usage. Chapter four discusses the implementation and results, focusing on the performance evaluation of the predictive models and the insights derived from the analysis. Chapter five explores engineering standards, societal impacts, and design challenges, addressing compliance with standards, sustainability, and complex engineering problem-solving methodologies. Finally, chapter six concludes the report by summarizing the findings, discussing the limitations, and providing directions for future work.

.

Chapter 2

Background

This chapter presents the impact of social media, its positive and negative aspects, and a review of relevant research. The chapter analyzes the impact of social media addiction and its related problems such as mental health, productivity, and relationships through various studies. In addition, a gap analysis identifies the limitations of previous research and suggests ways to overcome those limitations through our research.

2.1 Introduction

Social media is a digital platform that enables people to connect, share information and create all sorts of content. It is now very much indispensable in modern life and has had a considerable impact on individuals, society and businesses. Nevertheless, excessive or out of control usage has brought some problems such as addiction, mental health, productivity, and relationships. Machine learning is a kind of data-based decision-making method that can predict target based on data. So, we can use machine learning to predict the effect of people's lives on social media.

2.2 Literature Review

Akram, et al. [6] presented A Study on Positive and Negative Effects of Social Media on Society. Social media acts as a medium for enhancing public discussion and the exchange of viewpoints, although there exists a lacking in individuals' understanding about its true impacts. In this study they find positive sides of social networks as significant resources for professionals, showing of skills, provide access to business opportunities, and help with effective networking. Conversely, the internet carries inherent risks tied to online communities, such as cyberbullying—a form of harassment perpetrated using electronic technology.

Facebook and YouTube have become sources of addiction for Moghavvemi, S. et al. [7]: Malaysian students' usage patterns. The study investigates the utilization of Facebook and YouTube among Malaysian students, focusing on the potential for internet addiction and its impact on their social and academic lives. Using

Young's Internet addiction measure, 667 Facebook users and 1056 users of YouTube were surveyed for the study. According to the research, 22% of YouTube viewers and 18% of Facebook users are addicted, using these platforms for more than two hours every day.

A Survey on Facebook Addiction Level Among Selected Nigerian University Undergraduates was given by Alabi, O. F. [8]. The study examines the alarming issue of excessive Internet usage and its effects, which is commonly known as Internet Spectrum Addiction Disorder. The study concentrates specifically on Facebook Addiction as a possible manifestation of this disorder, with a particular emphasis on young adults in Nigeria. The purpose of this study is to determine the prevalence of Facebook addiction among selected Nigerian undergraduates. One thousand undergraduates from four universities in southwest Nigeria constituted the research sample. According to the study, there is a low prevalence of addiction (1.6%), especially among undergraduate students who attend private universities.

Gonzalez-Padilla, et al. [9] study the Social media influence ´ in the COVID-19 Pandemic. The COVID-19 pandemic has showcased the unprecedented speed of communication through social media platforms, playing a vital role in disseminating information. This study discussed the advantages and disadvantages of using social media during a pandemic. Also, the study suggests adhering to guidelines for sharing accurate and verified information.

The Impact of Social Media on Children, Adolescents, and Families was presented by G. S. O'Keeffe et al. [10]. The widespread use of social media among children and adolescents is a prevalent activity in today's digital landscape. The platforms serve as way for entertainment and communication for young individuals, leading to significant growth in their popularity. Hence, it becomes crucial for parents to understand the nature of these sites, as not all of them provide a healthy environment for children and teenagers.

Hajli, M., et al. [11] work on a study of the impact of social media on consumers. The study investigates how consumers use social media for online interaction and content creation, benefiting businesses. It validates a model using SEM-PLS, showing how social media transforms e-commerce into social commerce. Survey data reveals that social media enhances consumer interaction, building trust and purchasing intent. Trust and perceived website usefulness directly influence purchase intent. The paper concludes with discussions on results, implications, limitations, and future research directions.

Time Spent Using Social Media Impacts Mental Health: An Eight-Year Longitudinal Study, presented by Coyne, S. M., et al. [12]. Prior research has shown a connection between social media usage and mental health issues, but it is restricted by its cross-sectional design and absence of individual change analysis. Surprisingly, an 8-year study of 500 teenagers that followed their use of social media and mental health from age 13 to age 20 found no association

between social media usage and mental health problems. This refutes the idea that increased usage causes issues and promotes a wider perspective in mental health research that goes beyond screen time.

The work of Jafarkarimi, H., et al. [13] examines Facebook addiction in students from Malaysia. The study analyzed Facebook addiction among 441 Universiti Technologi Malaysia students using the Bergen Facebook Addiction Scale. Approximately 47% of participants showed addiction tendencies, which were consistent across postgraduates, undergraduates, and Malaysian and non-Malaysian students. Factors like religion, income, ego strength, and locus of control didn't significantly affect addiction risk. The research indicates a higher likelihood of addiction as Facebook usage increases.

Social media presents both benefits and challenges, according to Kaplan, A. M. et al. [14]. Social Media has emerged as an area of priority for organizations that wish to capitalize on sites like Wikipedia, YouTube, Face book and Tweeter for business gains. Nonetheless, its definition is still unclear and that is why this article exists to explain what Social Media is and how it is different from Web 2.0 and User Generated Content. It divides Social Media into six categories namely collaborative projects, blogs, content communities, social networking sites, virtual game worlds and virtual social worlds and provides ten guidelines to firms implementing Social Media.

Carr, C. et al. [15] discusses the absence of a clear definition of Social Media within communication scholarship and introduces a clear but versatile definition from numerous disciplines. It is informed by the growing trends in the pervasiveness and role of social media features, uses and users as it exposes theoretical hurdles in the future advancement of communication theories. The article focuses on how isolating and experimenting through the use of the social media benefits the understanding of both the Human-Computer Interface as well as the Human-Human interface and presents the framework for the future communication research.

Schober, M. F et al. [16] discusses if the information extracted from the social media messages can be used in addition to or as a replacement for survey data and discusses issues regarding reliability of such data for statistical purposes. It highlights the difference between survey researchers and data scientists in terms of sample representation, participant understanding and data validity. Hence, although social media content may not have the conventional population coverage, the success rate of predicting social phenomena is relative to the characteristics of the research problem, population under study, and/or the mode of analysis. Thus, close cooperation of these fields is necessary to overcome these problems.

Nti, I. K., Akyeramfo-Sam et al. [17] presented a study on Prediction of social media effects on students' academic performance using Machine Learning Algorithms (MLAs). The study investigated the relationship between students'

academic performance and the type of social networking site they use, how often they use it, how they use it in class, and how much exposure they have to it. They use Decision Tree (DT) and Random Forest (RF) machine learning algorithms to predicted the student's GPA. They observed that what students use SNSs for highly affect their academic performance; of the 100 participants, only 14.42% use SNSs for academic work, while 73.27% use SNSs for socialization (chatting). The study's conclusion demonstrates that students' academic performance is somewhat impacted by the frequency of social networking site use and its application in the classroom.

Table 2.1: Comparative Analysis

SL.	Year	Author	Objects (deal with)	Problem Domain	Size of Data	Algorithm	Best Accuracy
1	2022	Akter, M., Ritu, K. F. et al. [18]	Social Media Addiction	Prediction	504	Logistic Regression	94.05%
2	2019	Al Mamun, M. A., el al. [19]	Facebook addiction	Prediction	300	Regression analysis	39.7%
3	2024	Mim, M. N., Firoz et al. [20]	Social Media Addiction	Analysis	1417	Random Forest	82%
4	2020	Rahardjo, W., et al. [21]	Instagram addiction	Perception	434(Total) 259(Dtype)	Snowball sampling and regression techniques	56.9% (affected)
5	2022	Savci, M., Tekin, A. et al. [22]	Problematic social media	Prediction	309	SVM	63%
6	2024	This Work	Social Media Addiction	Prediction and Analysis	522	Random Forest Extra Trees LDA AdaBoosting Gaussian NB Stacking	91.71%

2.3 Gap Analysis

Many studies have examined the impact of social media addiction and how it can be analyzed and quantified, but there are notable gaps in the literature. Meanwhile, the vast majority of the existing studies have been focused on some specific social media platforms, e.g., Facebook or YouTube. Also, some studies only worked on either positive impact or negative impact. Furthermore, there is a lack of practical applicability, since most studies have been conducted at the theoretical level, with no useful tools or apps produced for users. To mitigate these gaps, our study considers nine social media platforms and talks about both the positive and negative impact of them. Also analyzed user behavior or daily activities based on data. In addition, a user-friendly mobile application has been developed, which collects data and provides addiction prediction. Our work overcomes previous limitations by using a diverse dataset and provides practical solutions.

2.4 Summary

Social media has a significant role in modern life, influencing people, society, and businesses in significant ways. Although excessive use has resulted in issues like addiction, mental health issues, and decreased productivity, it has also expanded chances for communication and information sharing. Prior research has typically been biased in its analysis of positive or negative effects and has been limited to particular social media sites. The majority of research has stayed theoretical in nature and has not produced any practical tools or apps for users. By examining nine platforms and evaluating both their advantages and disadvantages, our analysis gets over these restrictions. The daily behavior of users has been examined using a variety of datasets. Additionally, an easy-to-use mobile application that gathers information and offers addiction predictions, leading to practical solutions.

Chapter 3

Research Methodology

This chapter goes into detail on the research approach, framework, and procedure. A step-by-step presentation of the data collection, pre-processing, model selection, training, and evaluation processes is provided. The usefulness of several machine learning algorithms and ensemble techniques is explained, which contributed significantly to the research's accuracy and efficiency. Through the Flask API and Android application integration, users are provided with a useful tool that helps analyze social media addiction. The results of this research will help users increase awareness and develop healthy digital habits.

3.1 Methodology/Requirement Analysis & Design Specification

3.1.1 Overview

The steps involved in our study approach are highlighted in the following figure, Figure. 3.1. First of all, we have collected data through the Google form. Subsequently, we perform data preprocessing in consideration of several techniques, such as removing the features that bear low impacts on the dataset and employing encoding techniques to our dataset. Finally, the dataset is ready to fit the model. After that, RandomForestClassifier, ExtraTreesClassifier, LinearDiscrimantAnalysis, GaussianNB and models are established and trained. Since then, we attempted to enhance the model's performance through the use of Ensemble method. The Chosen models were selected based on higher accuracy. Across the whole process, the RandomForestClassifier, ExtraTreesClassifier, LinearDiscriminantAnalysis, GaussianNB and Ensemble learning model were being used.

3.1.2 Proposed Methodology/ System Design

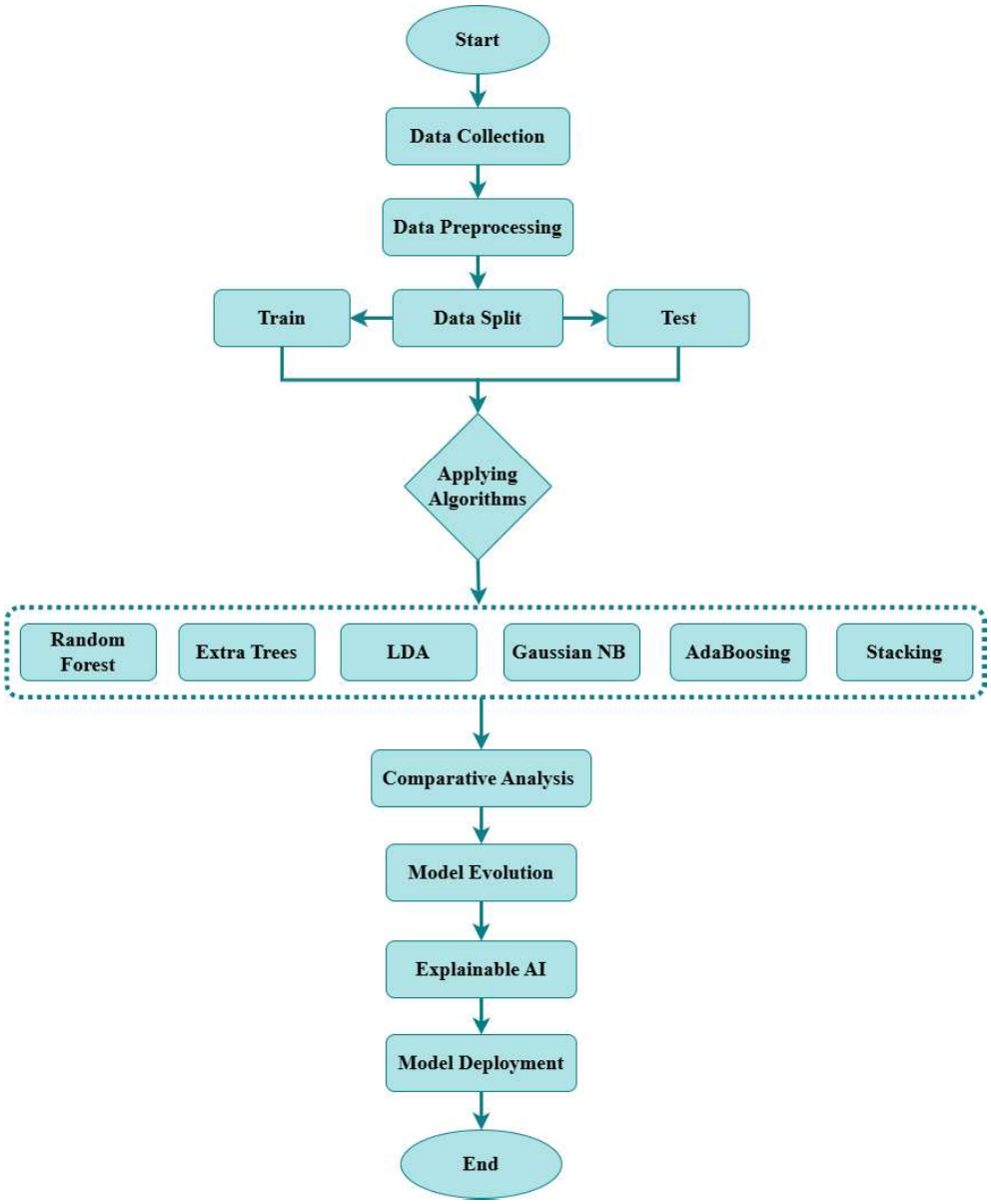

Figure 3.1: Execution Process

3.1.3 UI Design

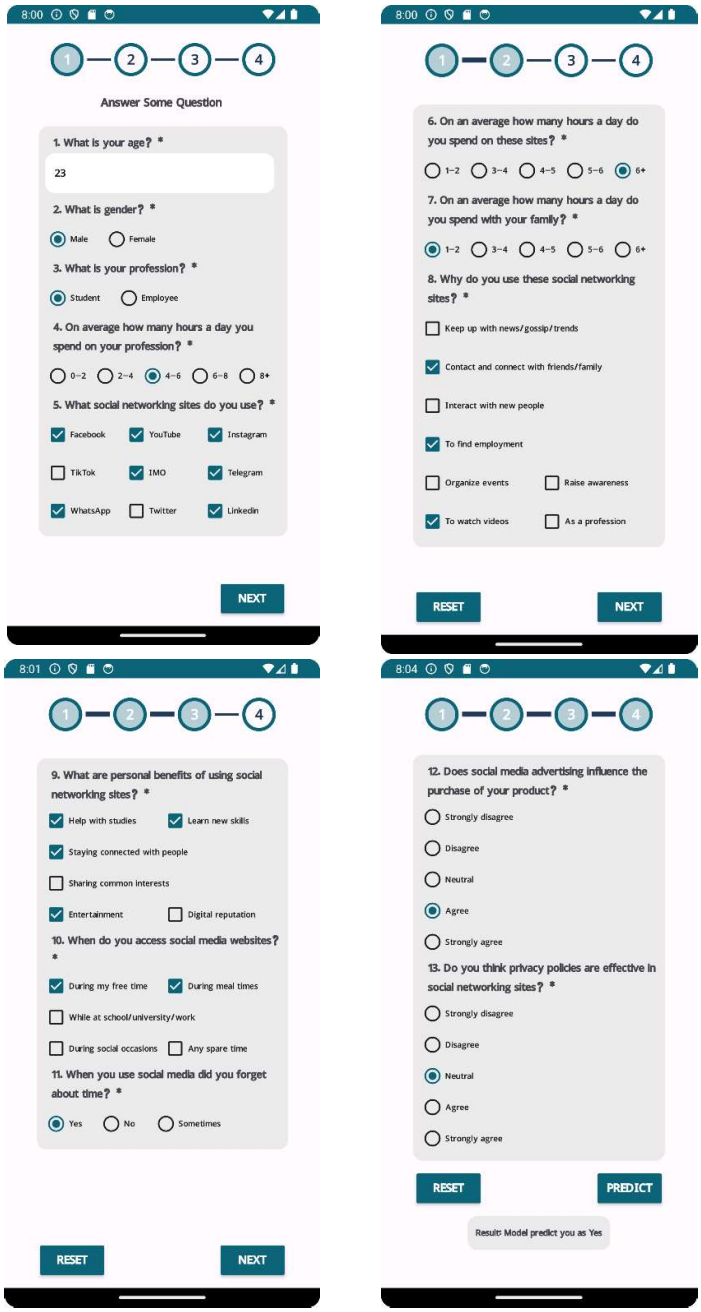

Figure 3.2: UI of Mobile Application

3.2 Detailed Methodology and Design

3.2.1 Dataset Description

In this study, we are concerned with the effects of social media on students and employees. We are collecting 522 data points. Two types of data are found in this dataset: numeric and string. We are students, so we can reach the student easily that's why most of the responses were generated from students, while an

additional few were from employees. Data was collected via a Google Form that was created to elicit as much information as possible that would be useful in the study. To ensure a thorough and representative dataset, we created a questionnaire comprising 14 questions. These questions were carefully chosen from published article and credible websites including SurveyMonkey, QuestionPro and Jotform. This approach helped to use the common formats of questions and guarantee the reliability and validity of the data.

Table 3.1: Description of Dataset

Age	13 - 19	5.75%
	20 - 30	80.27%
	30+	13.98%
Gender	Male	56.13%
	Female	43.87%
Profession	Student	78.16%
	Employee	21.84%
Time spends on profession	0 - 2	8.81%
	2 - 4	14.18%
	4 - 6	23.75%
	6 - 8	25.48%
	8+	27.78%
Social network site use	Facebook	79.31% of 100%
	YouTube	73.95% of 100%
	Instagram	45.02% of 100%
	TikTok	26.25% of 100%
	Imo	29.89% of 100%
	Telegram	42.15% of 100%
	WhatsApp	62.07% of 100%
	Twitter (X)	17.62% of 100%
	LinkedIn	20.69% of 100%
	Others	0.19% of 100%
Time spends on site	1 - 2	14.18%
	3 - 4	33.52%
	4 - 5	22.61%
	5 - 6	15.52%
	6+	14.18%
Time spends with family	1 - 2	32.18%
	3 - 4	24.52%
	4 - 5	19.16%
	5 - 6	9.77%
	6+	14.37%
Why use these social networking sites	keep up with news/gossip/ trends	55.8% of 100%
	contact and connect with friends/ family	73.4% of 100%
	interact with new people	32.5% of 100%
	to find employment	25.3% of 100%
	organize events	21.4% of 100%
	raise awareness	21.9% of 100%

	to watch videos	57.6% of 100%
	as a profession	14.7% of 100%
	Others	2.1% of 100%
Personal benefits of using social networking sites	help with studies	42.2% of 100%
	staying connected with people	69.3% of 100%
	learn new skills	49% of 100%
	entertainment	72.9% of 100%
	sharing common interests	27.8% of 100%
	digital reputation	17.2% of 100%
	Others	1.7% of 100%
When access social media websites	during my free time	76.5% of 100%
	while at school/university/work	46% of 100%
	during social occasions	31.8% of 100%
	during meal times	38.4% of 100%
	any spare time	28.7% of 100%
Privacy policies are effective	Strongly disagree	5.75%
	Disagree	8.62%
	Neutral	32.38%
	Agree	39.85%
	Strongly agree	13.41%
Advertising influences the purchase	Strongly disagree	4.60%
	Disagree	6.32%
	Neutral	28.35%
	Agree	39.85%
	Strongly agree	20.88%
Forget about time	Yes	41%
	No	29.89%
	Sometimes	29.11%
Addicted	Yes	42.91%
	No	33.14%
	Not Sure	23.95%

3.2.2 Preprocessing

In order to prepare raw data for model training, data preprocessing is an essential step in the machine learning workflow for building reliable, accurate, and efficient machine learning systems. It is necessary to appropriately preprocess the target dataset before any valuable information can be found [23]. We collect data through the Google form so, didn't have any null values in our dataset. We applied Encoding technique that convert the categorical variables into numerical values because machines can't understand without numerical value. Our dataset has nominal categories data that don't have rank order. Therefore, One-hot encoding is appropriate but this encoding in such cases results into very high dimensional vector representation which can pose a challenge to the storage and computations of machine learning models [24]. For this reason, Label Encoding is more appropriate than others to complete our task.

3.2.2.1 Label Encoding

Label encoding is the conversion of categorical data sets into a set of numbers. This transformation is particularly important when using a machine learning model because most of the algorithms accept only number inputs [25]. In label encoding, the specific column transforms it into numerical elements.
The label encoding assigns each category a unique integer:

$$\text{Encoded Value} = \begin{cases} 0, & for\ Cat_1 \\ 1, & for\ Cat_2 \\ & . \\ & . \\ & . \\ n-1, & for\ Cat_n \end{cases}$$

The encoding might be:

no	0
not sure	1
yes	2

3.2.3 Model Description

3.2.3.1 Random Forest

The random forest is a type of ensemble learning method using decision trees. It is one of the most powerful and popular machine learning algorithms for classification and regression tasks. Random Forests help to overcome some of the limitations of single decision trees and tend to result in higher predictive accuracy and robustness [26].
To use for Regression problems solving using Random Forest, then we can use mean squared error (MSE).

$$\text{MSE} = \frac{1}{N}\sum_{i=0}^{n}(f_i - y_i)^2 \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Were,

N - the number of points

f_i - The value returned by the model.

y_i - the actual value of the data point i

It helps determine which branch is best for the forest by calculating each node's distance from the expected real value. where fi is the result that the decision tree returns and y_i is the value of the data point that you are testing at a certain node.

If we are solving Classification problems using Random Forest, then Gini Index, or the Equation used to find how a node of a Decision Tree should branch out.

$$\text{Gini} = 1 - \sum_{i=1}^{c} (P_i)^2 \ldots\ldots\ldots\ldots\ldots (2)$$

Were,

Gini impurity/Gini index/Gini coefficient: A measure of impurity or disorder that is used by decision tree algorithms, especially in classification problems.

c is the number of unique classes or categories in the response variable.

P_i is the fraction of examples of class i in a node or a group of nodes.

This formula determines the Gini of each branch on a node and determines which branch is most likely to occur by using the class and probability. where c is the number of classes in the dataset and pi is the relative frequency of the class you are examining.

3.2.3.2 Extra Trees

ExtraTreesClassifier is an ensemble of learning that has its roots in decision tree learning. Extra-Trees Classifier is used for classification problems because of its great strength. This model is further bolstered by the Gorilla Troops Optimizer (GTO) and Reptile Search Algorithm (RSA), as are modern techniques used for optimizing decisions and increasing the accuracy of predictions [27]. These improvements allow the model to search for parameters spaces extensively and effectively in order to get the best parameter tuning. Thus, it outperforms a host of traditional tree-based classifiers across the varying dataset.

$$\hat{y} = \text{mode}(T_1(x), T_2(x) \ldots\ldots T_M(x)) \ldots\ldots\ldots\ldots\ldots (3)$$

Were,

Tm(x) is the prediction of the m-th tree for input x.

M is the total number of trees.

mode represents the majority vote across all tree predictions.

$$\text{Impurity}(S) = -\sum_{i=1}^{C} P_i \log(P_i) \ldots\ldots\ldots\ldots (4)$$

Were,

S is the dataset at a node.

C is the number of classes.

P_i is the proportion of class i in s.

The split reduces impurity by maximizing the information gain:

$$\Delta I = I(S) - \left(\frac{|S_L|}{|S|} I(S_L) + \frac{|S_R|}{|S|} I(S_R)\right) \ldots\ldots\ldots\ldots\ldots (5)$$

Were, S_L and S_R are the left and right child nodes after the split.

3.2.3.3 Linear Discrimination Analysis

LDA is simple linear classifier that allows us to reduce the dimensions of the data by projecting a dataset onto a lower-dimensional space through maximizing the class separability. An optimal discrimination projection matrix is defined as

the equitation (2) [28].

$$D_{otp} = \arg\max \frac{D^T S_B D}{D_T S_W D} = [d_1, d_2, \ldots . d_t]^T \ldots\ldots\ldots\ldots\ldots (6)$$

Were,

D_{opt} is optimal projection matrix.

S_B is between class scatter matrix.

S_W is within class scatter matrix.

Class-dependent and class-independent LDA techniques are the two categories of LDA techniques used to deal with classes. Each class has its own lower dimensional space that is computed for the purpose of projecting its data in the class-dependent LDA. In contrast, each class will be viewed as distinct from the others in the class-independent LDA. All classes can project their data onto a single lower-dimensional space of this type [29].

3.2.3.4 Gaussian NB

Gaussian NB is based on the Bayes theorem that supports continuous data derived from Gaussian normal distribution. NB is based on the assumption of independence among features. Assuming a normal distribution of data, this classifier is one of the simplest and more easily implementable supervised machine learning classification technique. The Bayes theorem is based on the assumption that each selected feature contributes to the decision equally and independently, which leads to multiply the likelihood and prior by the evidence. It assumes that the contribution of each feature to relevance is independent of its contributions to the others. Given that another event has already happened, the GNB is the probability of a particular occurring [30].

The Bayes Theorem:

$$P(A|B) = \frac{P(B|A) \, . \, P(A)}{P(B)} \ldots\ldots\ldots (7)$$

Were,

P(A) is the probability of a certain event happening without considering any additional information.

P(B | A) represents the probability of observing some evidence given that the event has occurred.

P(B) is the overall probability of observing the evidence, regardless of whether the event has occurred or not.

3.2.3.5 Ensemble Technique

Ensemble approaches enhance prediction accuracy when compared to single classifiers on a dataset. In this study, we used two ensemble methods to increase the performance of classification algorithms. Machine learning classifiers findings are combined using the two most prominent ensemble techniques: Boosting (AdaBoosting classifier) and Stacking classifier.

3.2.3.5.1 Boosting Classifier

Another widely used ensemble methods is boosting classifier. It is used to develop a group of classifiers. In boosting methods, a set of classifiers is learned sequentially where the classifiers fit data and then examine the mistakes. Decision trees are pruned sequentially to fit from the data and with the aim to achieve better accuracy in every stage [31]. Boosting is used to transform weak classifiers into robust models.

3.2.3.5.2 Stacking Classifier

To increase prediction accuracy, the stacking ensemble model includes base classifiers and a meta-classifier, which uses a variety of learning methods such as SVMs, neural networks, and decision trees. The meta-classifier is trained using the predictions of the base classifiers, improve their combination. Stacking is commonly used in a variety of applications, ensuring diversity and superior performance [32].

3.2.4 Performance Matrix

3.2.4.1 Confusion Matrix

Confusion matrix is one of several evaluation metrics that measure the performance of classification models. N X N is a confusion matrix, where N is the number of predicted classes. It is a table that outlines the quantity of ground truth instances of a certain class in relation to the quantity of expected class instances [33]. The predicted labels are classified in one of four categories: a true positive (TP) identifies correctly identified positives, a true negative (TN) defines correctly identified negatives, a false positive (FP) is a negative which was incorrectly found as a positive and finally the same definition can apply to false negative (FN), where the instance which is actually positive being determined as negative [34]. There are several performance metrics which we can compute directly from the confusion matrix such as accuracy, precision, recall and f1-score etc.

3.2.4.2 Accuracy

Accuracy represented the percentage of accurately categorized instances in the collection of all instances [34]. The accuracy is expressed as follows in equation (8).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \dots\dots\dots\dots\dots (8)$$

3.2.4.3 Precision

The percentage of true positives among all positive cases is determined by precision [35]. The precision is expressed as follows in equation (9).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \dots\dots\dots\dots\dots\dots (9)$$

3.2.4.4 Recall

Recall evaluates when an instance is actually positive, how often it predicts positive [36]. The recall is expressed as follows in equation (10).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \dots\dots\dots\dots\dots\dots (10)$$

3.2.4.5 F1-Score

Through the combination of precision and recall value, it evaluates the model's total efficacy [35]. In case of a classification problem, the F1-Score is nothing but the harmonic mean of precision and recall values. The recall is expressed as follows in equation (11).

$$\text{F1} - \text{Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \dots\dots\dots\dots (11)$$

3.2.5 Explainable AI (XAI)

Some of the XAI methods are SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model Agnostic Explanations), Grad-CAM (Gradient-weighted Class Activation Mapping), and Grad-CAM++. But we only significantly applied the LIME technique in our study. LIME constructs models that are closest to the local one for giving explanations to particular cases. These following ways help it contribute to XAI: feature importance, local explanation, and visualization.

3.2.5.1 Lime

LIME (Local Interpretable Model Agnostic Explanations) is the method of providing explanation in the context of the decision made by interpreter model in context of individual samples. LIME approximates the local behavior of a complex model around a specific instance by training a simpler, interpretable model [37]. LIME identifies the most influential features for a given prediction, which gives a better understanding of the model's decision-making process.

3.3 Deployment

The deployment of our machine learning model focuses on integrating the finalized solution into a production environment, making it accessible for practical use. In the context of analyzing and predicting social media addiction, the primary objective is to provide an accessible and user-friendly tool that empowers individuals to assess their social media usage patterns effectively.

This study deploys an end-to-end system that integrates the predictive model

with a functional Android application. The model, trained using ensemble stacking for optimal accuracy, was prepared for deployment by saving it in Pickle (.pkl) format. This conversion ensures lightweight and seamless integration with other platforms.

3.3.1 Backend System Development

A Flask API was developed to serve as the backend for the system. There are total 13 input parameters received by the API, which are features derived from our dataset, the model then processes this input through trained model and predicts whether the person is addicted to social media or not. The API is deployed on cPanel which can be accessed securely and can scale easily.

3.3.2 Android Application Integration

This Android application was created as a main user interaction interface. It contains a form with 13 questions corresponding to the features used in model training. Users feed their response, and this response is sent to deployed Flask API for analysis in real-time. The result of the API response on the app that represent the probability of social media addictive behavior. Our open access mobile application is available for download on our GitHub repository [38].

3.3.3 Technical Implementation

The frontend of the Android application was built in XML, and the backend functionality for the input was developed in Java. The application makes HTTP requests to the Flask API in order to communicate with it for data processing and exchange. The mobile application's user interface and real-time prediction are shown in Figure 3.2.

3.3.4 Benefits and Impact

The deployment framework creates a bridge between machine learning research and the real world, enabling people to gain meaningful insights from their social media behavior. Such a process could raise awareness about social media addiction and help users practice healthy digital habits.

3.4 Task Allocation

Table 3.2: Task Allocation

Tasks	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Improve the accuracy																	
Finalize the model and Prepare Mobile Application																	
Prepare detailed documentation																	

Estimated Work Period	
Actual Work Period	

3.5 Summary

This chapter discusses the methodology and framework used for the research. Described every single step in the process: from data collection to the pre-process, then we explained how we trained the model. The performance of several machine learning models, including Random Forest, Extra Trees, Linear Discriminant Analysis, Gaussian NB, and ensemble techniques, is emphasized. Also discuss a variety of evaluation metrics such as precision, accuracy, recall, F1-score, and more, focusing on combined methods compatible with one another in order to optimize the model's effectiveness. In addition, the use of Explainable AI methods, especially LIME, has made the research decision process more transparent. Through the Flask API and Android application integration, users are provided with a useful tool that helps analyze social media addiction. The results of this research will help users increase awareness and develop healthy digital habits

Chapter 4

Implementation and Results

This chapter analyzes the implementation process of our research and its related results. The environment setup, data processing, model training and their performance evaluation as well as the use of social media and its impact are discussed in detail.

4.1 Environment Setup

To perform our implementation, we use a PC with the Windows 10 operating system, an AMD Ryzen 5 5600G processor, and 16 GB of RAM. We execute the model using the open-source platform Jupyter Notebook. The Scikit-Learn framework is utilized to put the model into action (refer to Table Y).

Table 4.1: Experimental Setup.

Process Name	S.N.	Action
Input	1	Collected 522 samples based on daily activities from the social media user, including 3 classes.
Environment	2	Jupyter Notebook.
	3	Import all necessary libraries and packages.
	4	Load the dataset.
Splitting dataset	5	Split data training (70%) and testing (30%).
Training and Testing	6	Regression and Classification models are used for training on the train dataset. later, unseen data is given to predicts.
	7	For improving the predicts accuracy used Ensemble learning techniques.
Performance Evaluation	8	Generate model's accuracy.
	9	Generate classification reports.
	10	Generate confusion matrices.
Prediction	11	Load best model
	12	Load random sample
	13	Predict the classes

4.2 Testing and Evaluation/Performance/ Comparative Analysis

In our study, we applied six machine learning models to analyze and predict the impact of social media on daily life. The models are Random Forest, Extra Trees Classifier, Linear Discriminant Analysis (LDA), Gaussian Naive Bayes, Ensemble Technique's AdaBoost and Stacking. We used key metrics such as accuracy, precision, recall, and F1-score to evaluate the performance of these models. Confusion matrices Fig 4.1 and classification reports Table 4.2 for each model helped to analyze the performance of the models.

Table 4.2: Classification report of all model.

Model Name	Precision	Recall	F1-Score	Overall Accuracy
Random Forest	0.90	0.90	0.90	0.90
Extra Trees	0.91	0.90	0.90	0.90
LDA	0.91	0.90	0.90	0.90
Gaussian NB	0.89	0.89	0.89	0.89
AdaBoost	0.92	0.91	0.91	0.91
Stacking	0.92	0.92	0.92	0.92

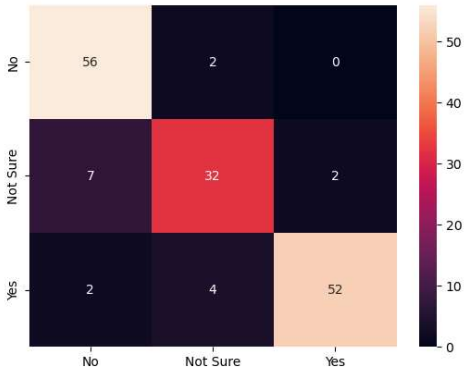

Figure 4.1(a): Confusion Matrix of Gaussian NB

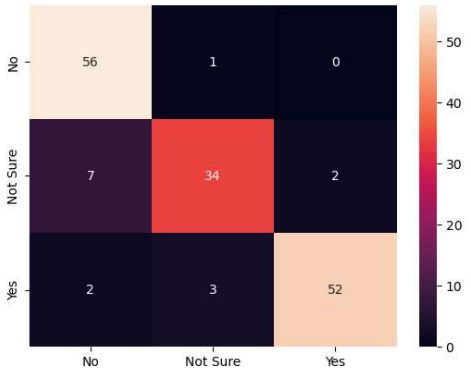

Figure 4.1(b): Confusion Matrix of LDA

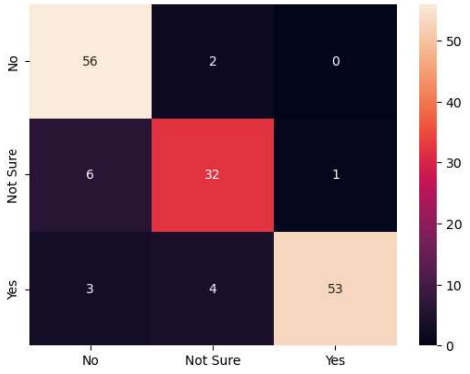

Figure 4.1(c): Confusion Matrix of Random Forest

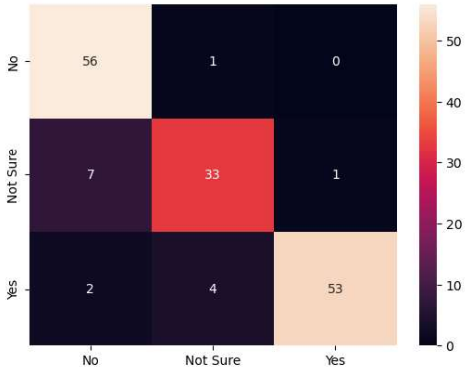

Figure 4.1(d): Confusion Matrix of Extra Trees

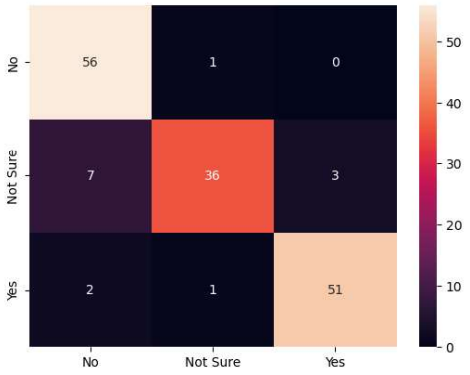

Figure 4.1(e): Confusion Matrix of AdaBoost

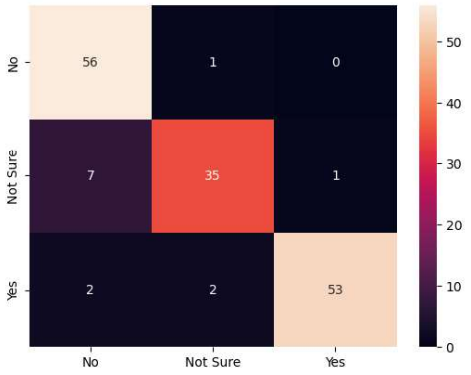

Figure 4.1(f): Confusion Matrix of Stacking

Figure 4.1: Confusion Matrix of all Model

The Gaussian NB model achieved 89.17% accuracy. Figure 4.1(a) represents the confusion matrix of the model. Among 157 test data. It correctly predicts 56 data as no, 32 data as not sure and 52 data as yes. And misclassified 17 data with each other. On the other hand, Linear Discriminant Analysis (LDA) model achieved 90.44% accuracy on the same dataset. In Figure 4.1(b) we can see that, how well it identifies the classes. In class no, the number of data occurrence is 57. It predicts 56 data correctly and incorrectly classified 1 data as not sure. In not sure class, the number of data occurrence is 43. It predicts 34 data correctly and misclassified 9 data with no and yes. In class yes, the number of data occurrence is 57. It correctly predicts 52 data. And misclassified 3 data with no and not sure.

Random Forest, Extra Trees Classifier and AdaBoost models achieved the same accuracy that is 91.08%. But there has some difference in class identification. Figure 4.1(c) represent the confusion matrix of Random Forest. The Random Forest model performed well in analyzing complex data relationships, and its ensemble method helped prevent overfitting. However, it made some misclassifications in certain classes. Where in class no, the number of data occurrence is 58. It predicts 56 data correctly and incorrectly classified 2 data as not sure. In not sure class, the number of data occurrence is 39. It predicts 32 data correctly and misclassified 7 data with no and yes. In class yes, the number of data occurrence is 60. It correctly predicts 53 data. And misclassified 7 data with no and not sure. While Figure 4.1(d) represent the confusion matrix of Extra Trees Classifier. In no class, among 57 data, the model correctly identifies 56 data and misclassified 1 data with not sure. In not sure class, it predicts 33 data correctly and misclassified 8 data with no and yes. In yes class, it predicts 53 correctly and misclassified 6 data with no and not sure class. On the other hand, Figure 4.1(e) represent the confusion matrix of AdaBoost. We can see that, how well it identifies the classes. in class no, the number of data occurrence is 57. It predicts 56 data correctly and incorrectly classified 1 data as not sure. In not sure class, the number of data occurrence is 46. It predicts 36 data correctly and

misclassified 10 data with no and yes. In class yes, the number of data occurrence is 54. It correctly predicts 51 data. And misclassified 3 data with no and not sure.

Among each model, the stacking ensemble technique performed best. Which achieved 91.71% accuracy. Stacking combined the predictions of the base models to take advantage of their complementary features. The meta model used in the stacking framework was able to analyze different patterns well, resulting in the highest accuracy and classification index. In Figure 4.1(f) highlights confusion matrix. Were in no class, it correctly classifies 56 data as no and misclassified 1 data with not sure. In class not sure, it predicts 35 data correctly and incorrectly classified 8 data with no and yes. In class yes, it identifies 53 data correctly. And misclassified 4 data with no and not sure.

Figure 4.2 shows the Overall performance of these model. Where stacking achieved the highest accuracy, which was 91.71%, and it was 91.08% better than the best individual model. And Gaussian NB achieved the low accuracy, that is 89.17%.

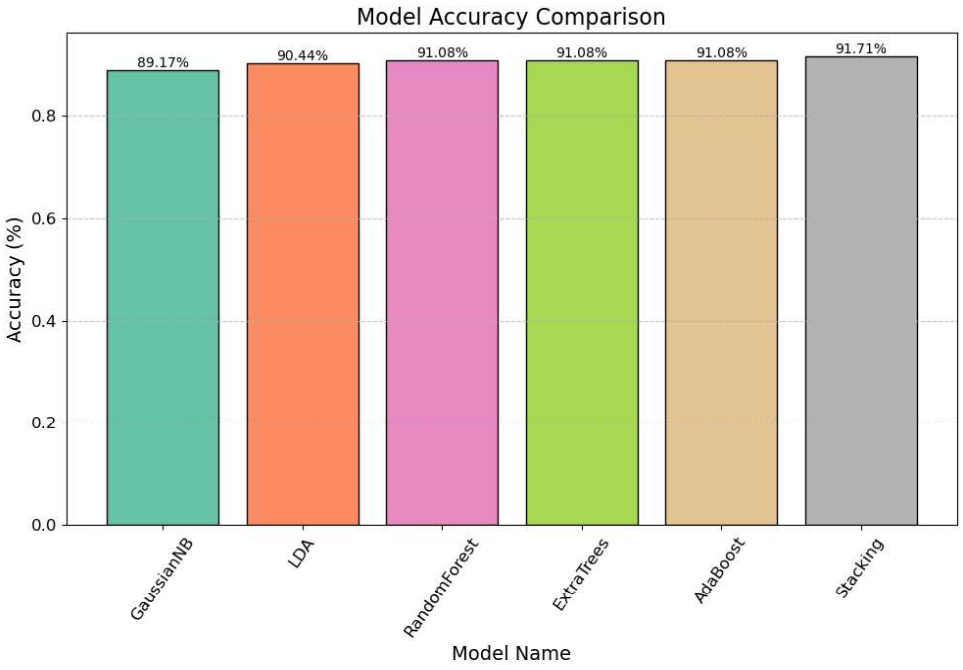

Figure 4.2: Model Comparison

4.3 Results and Discussion

There is no doubt that social media has quite a large contribution in today's society, which was primarily invented to help the people in society. On the other hand, recent studies have shown that social media abuse has grown to be a significant social problem. Its overuse and abuse can lead to detrimental effects on individuals' well-being and daily activities. Although it provides a number of benefits, like improved communication and information sharing.

Social media has completely changed the way we share, communicate, and

interact. It has enhanced our capacity to communicate with loved ones, have access to a plethora of knowledge, and interact with groups of people who share our interests. It's an excellent platform for professionals to network and advance their careers. Its reach for marketing and consumer involvement benefits businesses. Social media is essential for fostering social activity and increasing public awareness of social concerns. Additionally, it creates cross- cultural interactions, offers educational possibilities, and gives a feeling of support and belonging [39]. Social media, when utilized wisely, has the potential to have significant beneficial societal, personal, and professional effects.

Social media has a number of adverse effects in addition to its numerous positive impacts. Overuse can negatively affect relationships, productivity, and mental health. Particularly when users engage in excessive comparison or encounter cyberbullying, it may add to feelings of loneliness, anxiety, and despair. There are many privacy issues and data breaches, which raises concerns about the protection of personal data [40]. Furthermore, social media has the potential to be addicting, resulting in excessive screen time and decreased productivity.

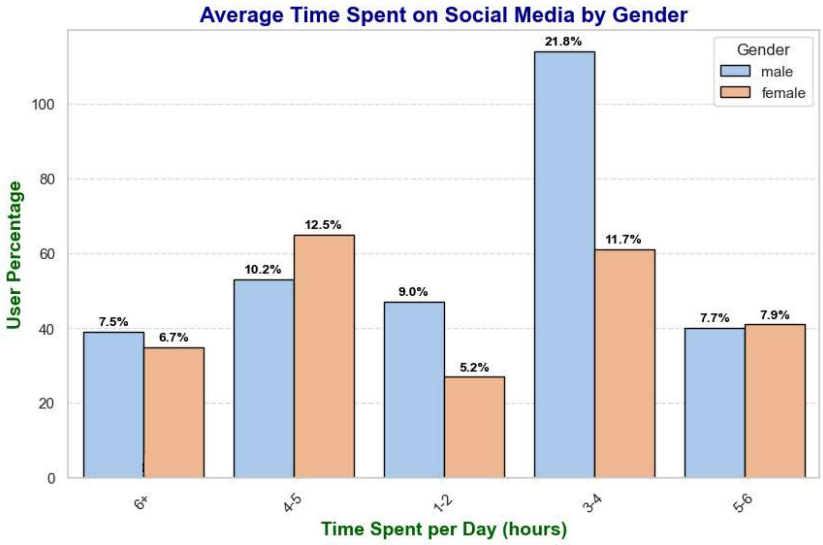

Figure 4.3: Time Spent on Social Media Across Gender

In Figure 4.3, the daily time spent on social media is shown by gender in various ranges. Males dominate a significant percentage of 3-4 hours users, 21.8%. It is an indicator of their proclivity for excessive use. Likewise, females show a slightly higher percentage in the 4-5 hours category compared to male counterparts, 12.5% and 10.2%, respectively, proving again that females have more mid-level engagement. Both genders have a similar distribution in the 5-6 hours range; however, the involvement of a large number of the population in the 6+ hours range may suggest overuse in both genders. These trends highlight the gender specificity in social media usage, with males preferring more intense but brief usage while females preferring longer-lasting sessions with the site.

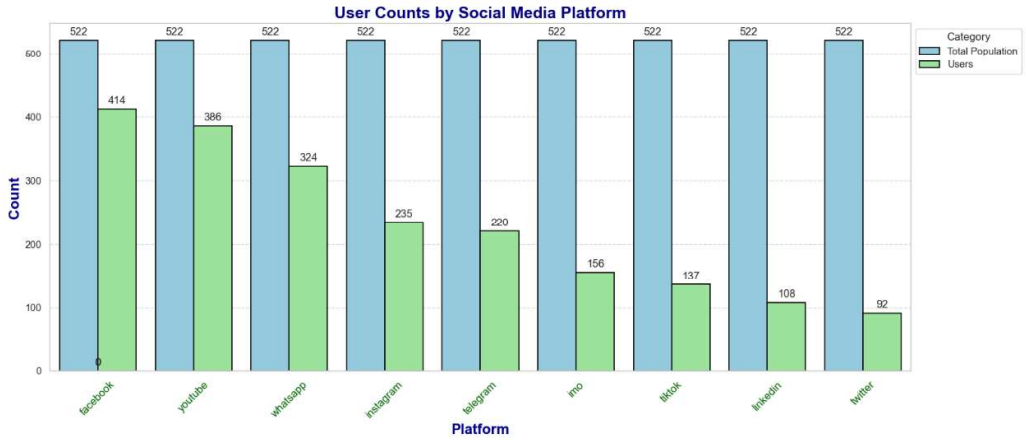

Figure 4.4: Social Media Uses Engagement

Figure 4.4 shows active social media platform use by 522 participants from the survey data where more than one social media platform could be chosen. Platform is indicated along the x-axis while respondent count is indicated along the y-axis. Two categories are visualized: Total Population, which remained the same for all the platforms, being 522, and Users, who referred to those who were actively using the various platforms. It is observed that the social media that has received more engagement from the users is Facebook, YouTube, and WhatsApp, while a lesser number of people actively use the media including Twitter, LinkedIn, TikTok, and other similar media. Such variation is due to the differences in the frequency of people's platform usage and activity levels within the survey audience. Such constancy of the total population value aids in uniformity across various platforms, which provides a proper framework for analysis of the use and preference for social media platforms. In particular, this visualization assists the research by providing clear insight into the platform's engagement trends.

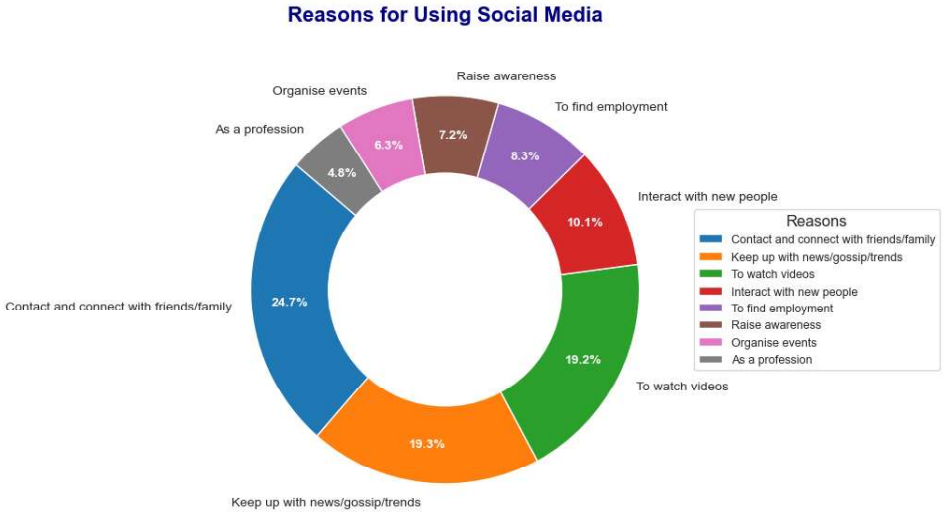

Figure 4.5: Reason for Using Social Media

The Figure 4.5 shows the reasons why people use social media, based on survey responses in which participants could select several options. The most common reason, chosen by 24.7% of participants, is to contact and connect with friends and family, demonstrating the importance of social media in maintaining relationships. This is followed by getting up to date on news, gossip, and trends (19.3%) and watching videos (19.2%), emphasizing its dual role as an information and entertainment provider. Interacting with new people (10.1%) and increasing awareness (8.3%) indicate that social media is also useful for extending social circles and advocacy. Professional usage, such as job seeking (7.2%), organizing events (6.3%), and professional use of social media (4.8%), are less common, indicating the platform's predominantly personal and recreational nature. While social media serves multiple purposes, its primary interest is in connecting people and providing entertainment.

4.4 Summary

In this chapter, the implementation and results of our research are analyzed. The chapter initially discusses the process of setting up the environment and using the dataset. Then, the performance of different machine learning models is analyzed where the stacking method performed best. Next, the use of social media and its impact are discussed in detail, where the behavior of users, daily time spent, and platform usage trends are analyzed.

Chapter 5

Engineering Standards and Design Challenges

This chapter discusses in detail the engineering criteria used in the project, the design challenges, and the impact of the project on society and the environment. The project uses software and hardware that meets international standards, and ethics, social development, and environmental aspects are given special importance.

5.1 Compliance with the Standards

5.1.1 Software Standards

Common industry standards have been followed to ensure the quality and functionality of the software systems used in our project. Scikit-Learn and TensorFlow libraries have been used for machine learning models, which are well-known for their stability and efficiency. Flask has been used for API development, which is reliable for data processing and client-server communication. Android SDK and Java have been used for mobile application development, which are proven and compatible for the Android platform.

5.1.2 Hardware Standards

The hardware used follows well-known standards to ensure maximum performance and reliability. A desktop with an AMD Ryzen 5 5600G processor and 16GB RAM was used for training and testing. Such hardware is powerful enough for research and model training. The cPanel used for API hosting provides reliable and scalable services, suitable for small to large data sets.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

Our project can have a positive impact on people's daily lives. Our mobile application for assessing social media addiction will make users aware of their

usage patterns. This will help protect their mental health and improve digital time management skills. Furthermore, it will help maintain a balance between personal and professional life, which can improve people's quality of life in the long run.

5.2.2 Impact on Society & Environment

Our project will play an important role in raising awareness in society. By raising awareness about social media addiction, it will help individuals and society develop healthy digital habits. It will have an impact on increasing the effectiveness of workers and students. Environmentally, the project provides a completely software-based solution without using any physical resources, which is environmentally friendly.

5.2.3 Ethical Aspects

Ethical aspects have been given special importance in this project. Complete confidentiality and data security have been ensured while collecting and processing user data. Transparency and fairness have been maintained in the use of data. Furthermore, the project in no way encourages users to use excessive technology, but rather encourages them to develop healthy digital habits by raising their awareness.

5.3 Project Management and Financial Analysis

Table 5.1: Financial Analysis.

SN	Components	Estimated Cost (BDT)
01.	Computer	120000-130000
02.	Collecting Data	500-1000
05.	Software	5000-6000
06.	Documentation and Report Writing	500-1000
07.	Contingency (10%of total)	1500-2000
	Total Estimated Cost	128500-140000

5.4 Complex Engineering Problem

This section outlines how the project satisfies complex engineering problems (EPs) and their corresponding Knowledge Profiles (K) and Engineering Activities (EA). The mappings for each EP are defined below, followed by a description.

5.4.1 Complex Problem Solving

Table 5.2: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓			✓	✓

5.4.1.1 Mapping with Knowledge Profile for EP1

The project required a deep understanding of machine learning fundamentals, ensemble techniques, and app deployment. Literature reviews enriched the foundation of specialist knowledge applied.
This table 5.3) is designed to map the EP1 to the Knowledge Profile.

Table 5.3: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓			✓

K3 (Engineering Fundamentals): The project required fundamental knowledge of machine learning algorithms, including Random Forest, Extra Trees, and ensemble models like AdaBoost and Stacking.

K4 (Specialist Knowledge): Advanced understanding of ensemble learning and Explainable AI techniques (LIME) was used to interpret model behavior.

K8 (Research Literature): Conducted a detailed literature review to identify gaps in previous studies on social media addiction and integrate insights.

5.4.1.2 Mapping with Knowledge Profile for EP2

Balancing accuracy, runtime performance, and user-friendliness posed conflicting requirements, resolved by optimizing the model and ensuring lightweight API responses.
This table 5.4) is designed to map the EP2 to the Knowledge Profile.

Table 5.4: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓		✓	✓	

K3 (Engineering Fundamentals): Balancing model accuracy (91.71% for Stacking) with computational efficiency and ensuring real-time response for app users.

K5 (Engineering Design): Designing a user-friendly Android application while ensuring data security and fast API response times.

K6 (Engineering Practice): Resolving trade-offs between prediction accuracy and system complexity to deploy the solution in a real-world environment.

5.4.1.3 Mapping with Knowledge Profile for EP3

Comprehensive analysis was performed by testing six models, fine-tuning hyperparameters, and employing advanced evaluation techniques.
This table 5.5) is designed to map the EP3 to the Knowledge Profile.

Table 5.5: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓			✓

K3 (Engineering Fundamentals): Applied preprocessing techniques such as label encoding and feature selection to prepare the dataset.

K4 (Specialist Knowledge): Evaluated six machine learning models on metrics like accuracy, precision, recall, and F1-score.

K8 (Research Literature): Used prior research to validate the methodology and justify the selection of ensemble techniques.

5.4.1.4 Mapping with Knowledge Profile for EP6

Stakeholders (students, employees) contributed through data collection, influencing the development of the predictive model.

This table 5.6) is designed to map the EP6 to the Knowledge Profile.

Table 5.6: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
			✓	

K6 (Engineering Practice): Collected data from stakeholders (students and

employees) via structured questionnaires, ensuring their input shaped the final model.

5.4.1.5 Mapping with Knowledge Profile for EP7

This interdisciplinary project combined data science, machine learning, and mobile app development to create a holistic solution.
This table 5.7) is designed to map the EP7 to the Knowledge Profile.

Table 5.7: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
		✓	✓	

K5 (Engineering Design): Combined machine learning and software engineering to develop an end-to-end solution.

K6 (Engineering Practice): Integrated machine learning predictions with a functional mobile app to bridge research and practical application.

5.4.2 Engineering Activities

Table 5.8: Mapping with complex engineering activities.

EA1 Range of re-sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓		✓	✓

5.4.2.1 EA1 Range of resource

Utilized Scikit-learn, TensorFlow, Flask, and Android Studio for model training, API development, and mobile application integration.

5.4.2.2 EA2 Level of Interaction

Engaged stakeholders (students and employees) for data collection and collaborated with supervisors to refine methodology.

5.4.2.3 EA4 Consequences for society and environment

Raised awareness of social media addiction, promoting healthier digital habits. Provided an environmentally friendly solution through a software-only approach.

5.4.2.4 EA5 Familiarity

Leveraged existing machine learning techniques while introducing a novel deployment framework to address familiar issues in a new way.

5.5 Summary

The chapter discusses the engineering standards and design challenges of the project. Quality has been ensured through the use of software and hardware following international standards. The project has had a positive impact on raising awareness in the society and improving digital health and has been implemented in an environmentally friendly manner. Data security and privacy have been ensured while maintaining ethics. Complex engineering problems have been solved through machine learning and app development, where an effective solution has been created through advanced algorithms and stakeholder participation.

Chapter 6

Conclusion

6.1 Summary

As technology advances, social media has become a fundamental part of everyday life. Social media's rapid growth has had significant effects on numerous areas of human life, including behavior, communication, productivity, and mental health. This study used machine learning techniques, specifically the StackingClassifier model, to look at and predict these impacts based on patterns of user engagement, sentiment, and activity. The findings demonstrate how machine learning models can effectively classify and predict the results of social media use, providing foundations for developing data-driven solutions to reduce negative outcomes. This study further demonstrates the significance of moral issues in handling user data and ensuring accurate model predictions. The future research needs to focus on expanding datasets that include a variety of demographic information, updating algorithms for more accuracy, and evaluating the long-term social media societal effects. In brief, social media persists as an effective tool for creation and connection, but its impact on day-to-day life needs to be carefully considered and controlled. Society may more effectively use machine learning's advantages while solving its weaknesses by using it with interdisciplinary approaches.

6.2 Limitation

One of the weaknesses of this study is that it collected a small amount of data for analysis. This study mainly focuses on students with fewer employed people and other groups included in the study, thus limiting the external validity of the results. Emphasis on only 9 well-known social networks does not include relatively new social platforms or local platforms that can have different effects. Using Google Forms to capture self-report data often has disadvantages such as social desirability bias and recall bias.

6.3 Future Work

By acquiring a wider and more diverse dataset, future research should try to overcome these limitations and ensure more representation across a range of populations, such as professionals and other minority groups. A more comprehensive understanding of the impact of social media will be possible if the focus goes beyond the 14 well-known social networks to include a variety of area various platforms. Biases like social desirability and recall bias can be reduced by using more reliable data gathering techniques, such as observational research or app-based tracking.

References

[1] S. Siddiqui and T. Singh, "Social media its impact with positive and negative aspects," International Journal of Computer Applications Technology and Research, vol. 5, no. 2, pp. 71-75, 2016.

[2] S. Kemp, "Social media statistics for Bangladesh in 2023," DataReportal, Feb. 13, 2023. [Online]. Available: https://datareportal.com/reports/digital -2023-bangladesh.

[3] S. Kemp, "Social media statistics for Bangladesh in 2024," DataReportal, Feb. 23, 2024. [Online]. Available: https://datareportal.com/reports/digital -2024-bangladesh.

[4] S. Kemp, "Social media users in Bangladesh," DataReportal, Feb. 17, 2020. [Online]. Available: https://datareportal.com/reports/digital-2020-bangaldesh.

[5] R. P. Care, "Clinical report—the impact of social media on children, adolescents, and families," Pediatrics, vol. 127, no. 4, p. 801, 2011.First Author and Second. Article title. *Journal Name*, Volume Number(Issue Num-ber):Start Page–End Page, Year of Publication.

[6] W. Akram and R. Kumar, "A study on positive and negative effects of social media on society," International Journal of Computer Sciences and Engineering, vol. 5, no. 10, pp. 351-354, 2017.

[7] S. Moghavvemi, A. B. Sulaiman, N. I. Jaafar, and N. Kasem, "Facebook and YouTube addiction: The usage pattern of Malaysian students," in 2017 International Conference on Research and Innovation in Information Systems (ICRIIS), 2017, pp. 1-6.

[8] O. F. Alabi, "A survey of Facebook addiction level among selected Nigerian University undergraduates," New Media and Mass Communication, vol. 10, pp. 70-80, 2013.

[9] D. A. Gonzalez-Padilla and L. Tortolero-Blanco, "Social media's influence in the COVID-19 Pandemic," International Braz J Urol, vol. 46, pp. 120-124, 2020.

[10] G. S. O'Keeffe and K. Clarke-Pearson, "The impact of social media on children, adolescents, and families," Pediatrics, vol. 127, no. 4, pp. 800-804, 2011.

[11] M. N. Hajli, "A study of the impact of social media on consumers," International Journal of Market Research, vol. 56, no. 3, pp. 387-404, 2014.

[12] S. M. Coyne, A. A. Rogers, J. D. Zurcher, L. Stockdale, and M. Booth, "Does time spent using social media impact mental health? An eight-year longitudinal study," Computers in Human Behavior, vol. 104, p. 106160, 2020.

[13] H. Jafarkarimi, A. T. H. Sim, R. Saadatdoost, and J. M. Hee, "Facebook addiction among Malaysian students," International Journal of Information and Education Technology, vol. 6, no. 6, pp. 465, 2016.

[14] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of social media," Business Horizons, vol. 53, no. 1, pp. 59-68, 2010.

[15] C. T. Carr and R. A. Hayes, "Social media: Defining, developing, and divining," Atlantic Journal of Communication, vol. 23, no. 1, pp. 46-65, 2015.

[16] M. F. Schober, J. Pasek, L. Guggenheim, C. Lampe, and F. G. Conrad, "Social media analyses for social measurement," Public Opinion Quarterly, vol. 80, no. 1, pp. 180-211, 2016.

[17] I. K. Nti, S. Akyeramfo-Sam, B. Bediako-Kyeremeh, and S. Agyemang, "Prediction of social media effects on students' academic performance using Machine Learning Algorithms (MLAs)," Journal of Computers in Education, vol. 9, no. 2, pp. 195-223, 2022.

[18] M. Akter, K. F. Ritu, M. T. Habib, M. S. Rahman, and F. Ahmed, "A Machine Learning Approach to Predict Social Media Addiction During COVID-19 Pandemic," in 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 401-405.

[19] M. A. Al Mamun and M. D. Griffiths, "The association between Facebook addiction and depression: A pilot survey study among Bangladeshi students," Psychiatry Research, vol. 271, pp. 628-633, 2019.

[20] M. N. Mim, M. Firoz, M. M. Islam, M. Hasan, and M. T. Habib, "A study on social media addiction analysis on the people of Bangladesh using machine learning algorithms," Bulletin of Electrical Engineering and Informatics, vol. 13, no. 5, pp. 3493-3502, 2024.

[21] W. Rahardjo and I. Mulyani, "Instagram addiction in teenagers: The role of type D personality, self-esteem, and fear of missing out," Psikohumaniora: Jurnal Penelitian Psikologi, vol. 5, no. 1, pp. 29–44, 2020.

[22] M. Savci, A. Tekin, and J. D. Elhai, "Prediction of problematic social media use (PSU) using machine learning approaches," Current Psychology, vol. 41, no. 5, pp. 2755-2764, 2022.

[23] S. Rao, P. Poojary, J. Somaiya, and P. Mahajan, "A comparative study

between various preprocessing techniques for machine learning," Int. J. Eng. Appl. Sci. Technol., vol. 5, no. 3, pp. 2455-2143, 2020.

[24] M. K. Dahouda and I. Joe, "A deep-learned embedding technique for categorical features encoding," IEEE Access, vol. 9, pp. 114381-114391, 2021.

[25] Kattilaxman, "A Practical Guide for Python: Label Encoding with Python," Medium, Oct. 25, 2023. [Online]. Available: https://medium.com/@kattilaxman4/a-practical-guide-for-python-label-encoding-with-python-fb0b0e7079c5.

[26] I. Reis, D. Baron, and S. Shahaf, "Probabilistic random forest: A machine learning algorithm for noisy data sets," The Astronomical Journal, vol. 157, no. 1, p. 16, 2018.

[27] Y. Li and M. He, "Elevating Student Performance Prediction using Extra-Trees Classifier and Meta-Heuristic Optimization Algorithms," International Journal of Advanced Computer Science & Applications, vol. 15, no. 2, 2024.

[28] J. Qi, P. Yang, A. Waraich, Z. Deng, Y. Zhao, and Y. Yang, "Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: A systematic review," Journal of Biomedical Informatics, vol. 87, pp. 138-153, 2018.

[29] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," AI Communications, vol. 30, no. 2, pp. 169-190, 2017.

[30] S. Naiem, A. E. Khedr, A. M. Idrees, and M. I. Marie, "Enhancing the efficiency of Gaussian naïve Bayes machine learning classifier in the detection of DDoS in cloud computing," IEEE Access, vol. 11, pp. 124597-124608, 2023.

[31] R. Singh and S. Pal, "Machine learning algorithms and ensemble technique to improve prediction of students' performance," International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, no. 3, 2020.

[32] M. Kumar, S. Singhal, S. Shekhar, B. Sharma, and G. Srivastava, "Optimized stacking ensemble learning model for breast cancer detection and classification using machine learning," Sustainability, vol. 14, no. 21, p. 13998, 2022.

[33] J. M. Ph.D. and E. Kavlakoglu, "What is a confusion matrix," IBM, 2024. [Online]. Available: https://www.ibm.com/topics/confusion-matrix.

[34] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," Scientific Reports, vol. 14, no. 1, p. 6086, 2024.

[35] O. Islam, M. Assaduzzaman, and M. Z. Hasan, "An explainable AI-based blood cell classification using optimized convolutional neural network," Journal of Pathology Informatics, vol. 15, p. 100389, 2024.

[36] N. Japkowicz, "Why question machine learning evaluation methods," in AAAI Workshop on Evaluation Methods for Machine Learning, vol. 6, no. 11, 2006.

[37] E. Mahamud, N. Fahad, M. Assaduzzaman, S. M. Zain, K. O. M. Goh, and M. K. Morol, "An explainable artificial intelligence model for multiple lung diseases classification from chest X-ray images using fine-tuned transfer learning," Decision Analytics Journal, vol. 12, p. 100499, 2024.

[38] M. A. Kayum and H. H. Tuhin, "SMI Prediction Mobile App," GitHub, 2024. [Online]. Available: https://github.com/mdabdulkayum/SMI_Prediction.

[39] G. Zaidi, "Positive effects of social media: The upside of staying connected," Social Champ, May 2, 2024. [Online]. Available: https://www.socialchamp.io/blog/positive-effects-of-social-media/.

[40] G. Raju, "Impact of Social Media: Advantages and Disadvantages [2024]," Simplilearn, Oct. 7, 2024. [Online]. Available: https://www.simplilearn.com/real-impact-social-media-article.

SMIDefenceTemplate.pdf

ORIGINALITY REPORT

25%	19%	12%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	5%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
3	Submitted to United International University Student Paper	1%
4	Eram Mahamud, Nafiz Fahad, Md Assaduzzaman, S.M. Zain, Kah Ong Michael Goh, Md. Kishor Morol. "An explainable artificial intelligence model for multiple lung diseases classification from chest X-ray images using fine-tuned transfer learning", Decision Analytics Journal, 2024 Publication	1%
5	www.researchgate.net Internet Source	1%
6	academic-accelerator.com Internet Source	<1%
7	www.mdpi.com Internet Source	<1%