VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Project Report On

"The Digital Mindscape: Leveraging Machine Learning To Understand Social Media's Effects on Human Mental Health"

In partial fulfillment of the Fourth Year (Semester–VII), Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2024-2025

Project Mentor: Ms. Mannat Doultani

Submitted by

Aryan Manghi D17B/28

Prasad Chaudhari D17B/08

Devyaansh Razdan D17B/45

(2024-25)

VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

CERTIFICATE of Approval

This is to certify that **Aryan Manghi (D17B/28), Prasad Chaudhari (D17B/08), Devyaansh Razdan (D17B/45)** of Fourth Year Computer Engineering studying under the University of Mumbai has satisfactorily presented the project on **"The Digital Mindscape: Leveraging Machine Learning to Understand Social Media's Effects on Human Mental Health."** as a part of the coursework of PROJECT-I for Semester-VII under the guidance of **Ms. Mannat Doultani** in the year 2024-2025.

Date

Internal Examiner

External Examiner

Project Mentor

Head of the Department
Dr. Mrs. Nupur Giri

Principal
Dr. J. M. Nair

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Ms. Mannat Doultani** for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to the Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times

Computer Engineering Department

COURSE OUTCOMES FOR B.E PROJECT

Learners will be to:-

Course Outcome	Description of the Course Outcome
CO 1	Do literature survey/industrial visit and identify the problem of the selected project topic.
CO2	Apply basic engineering fundamental in the domain of practical applications FORproblem identification, formulation and solution
CO 3	Attempt & Design a problem solution in a right approach to complex problems
CO 4	Cultivate the habit of working in a team
CO 5	Correlate the theoretical and experimental/simulations results and draw the proper inferences
CO 6	Demonstrate the knowledge, skills and attitudes of a professional engineer & Prepare report as per the standard guidelines.

Abstract of the Project

Social media plays a significant role in modern life, shaping how individuals connect and engage with others. However, its impact on mental health is complex, with both positive and negative effects. The "Impact of Social Media on Mental Health using ML" project aims to explore this relationship using machine learning to analyze large datasets from social media platforms and mental health surveys. By identifying patterns and correlations, the project seeks to uncover key insights into how social media behaviors influence mental well-being, providing evidence-based solutions to mitigate harm and enhance positive outcomes.

A core component of the project is the development of an intelligent chatbot, which offers personalized mental health recommendations based on individual user interactions. This chatbot, powered by machine learning, will analyze user input and social media engagement to provide tailored advice that addresses the unique mental health challenges faced by each user. This real-time feedback will empower users to make healthier decisions about their social media habits and mental health. By offering personalized support, the platform aims to reduce anxiety, stress, and other negative mental health outcomes associated with prolonged or harmful social media use. Ultimately, the project strives to create a balanced digital environment where users can enjoy the benefits of social media without compromising their mental health, contributing to a healthier digital society.

INDEX

Chapter 1: Introduction

1.1 Introduction

In the ever-evolving digital landscape, social media has become an integral part of everyday life, particularly for children and adolescents. While these platforms offer spaces for creativity, communication, and social interaction, they also come with a dark side that poses serious risks to mental health. The increasing prevalence of anxiety, depression, and social withdrawal in children has been linked to the pressures of maintaining online personas, exposure to curated and often unrealistic content, and the addictive nature of social media platforms. This project, "The Digital Mindscape: Leveraging Machine Learning to Understand Social Media's Effects on Child Mental Health," aims to explore and address these pressing concerns through advanced data analytics and machine learning techniques.

Our initiative seeks to uncover the intricate and potentially harmful connections between social media usage and mental health outcomes in children. By analyzing various factors, including screen time, content engagement, and interaction patterns, we aim to identify key risk factors that contribute to social media addiction and its negative mental health implications. Machine learning provides a powerful toolset to process and interpret these large datasets, revealing hidden patterns and triggers that may otherwise go unnoticed. These insights will serve as the foundation for developing strategies to mitigate the detrimental effects of social media on child mental health.

Ultimately, the goal of this project is to raise awareness and provide actionable solutions for both children and their guardians. By creating an accessible platform that delivers personalized insights, we hope to guide users toward healthier social media habits. In doing so, we aim to reduce the growing mental health issues that have become synonymous with excessive or inappropriate social media usage among younger demographics.

1.2 Motivation

The motivation behind this project stems from the alarming increase in mental health issues among children, coinciding with the rise of social media use. Research has shown a strong correlation between social media activity and various mental health challenges, including anxiety, depression, and diminished self-esteem. Children are particularly vulnerable to the negative effects of social media due to their developing minds and the pressure to conform to idealized images and lifestyles portrayed online. These platforms are specifically designed to maximize user engagement, often to the detriment of psychological well-being. The lack of awareness around the extent of this issue and the limited access to tools that allow users to manage their digital consumption motivated us to explore data-driven solutions.

Moreover, current measures to mitigate the harmful impact of social media are often generalized and do not address the nuanced ways in which individual users are affected. By leveraging machine learning to deeply analyze social media patterns, this project aims to fill that gap, offering personalized insights that can help children, parents, and educators make informed decisions. We are motivated by the potential to contribute meaningfully to this growing area of concern by providing scalable, effective interventions.

1.3 Problem Definition

In today's hyperconnected world, children and adolescents spend increasing amounts of time on social media, exposing them to a variety of digital content and interactions. While these platforms offer opportunities for social engagement and creativity, they also introduce significant mental health risks, including anxiety, depression, and social comparison. The complex ways in which screen time, types of content consumed, and interaction patterns affect a child's mental well-being are not fully understood. To address this, our project seeks to collect and analyze real-time data on children's social media usage, offering a more comprehensive and data-driven understanding of these relationships.

By leveraging advanced machine learning algorithms, we plan to uncover patterns and correlations within this vast dataset to identify how particular elements of social media—such as exposure to certain types of content (e.g., negative or harmful material), prolonged screen time, and addictive behaviors—contribute to mental health challenges. The project will focus on factors like the frequency and duration of social media use, the nature of user interactions (such as receiving likes or comments), and engagement with various types of content (positive vs. negative). This analysis will help pinpoint key risk factors that increase the likelihood of mental health issues. For example, we may discover that children who are more frequently exposed to curated, idealized portrayals of life are more likely to suffer from low self-esteem or anxiety due to constant social comparison.

The project aims to not only identify risks but also provide actionable insights to help users manage their social media habits. Through an interactive platform, users will track screen time, content engagement, and interactions. Machine learning algorithms will analyze this data to offer personalized feedback, such as reducing harmful content exposure and moderating screen time. A user-friendly dashboard will allow children and guardians to monitor progress and adjust behaviors.The platform will provide real-time alerts when risky patterns, like prolonged screen time or negative content exposure, are detected. Users will also receive tips for healthier social media use, promoting habits like mindful engagement and regular breaks.

1.4 Relevance

The relevance of this project lies in its direct response to the growing mental health crisis among children in the digital age. As social media becomes increasingly ingrained in the daily lives of young users, there is an urgent need to understand the potential psychological harm it may cause. With children spending significant portions of their day online, often exposed to unrealistic portrayals of life, curated content, and social comparison, the impact on their mental well-being cannot be understated. Existing interventions for mental health challenges related to social media use are limited, often reactive, and lack personalization. By using machine learning to analyze real-time data, this project offers a proactive, data-driven approach to identifying specific risks associated with social media usage, helping children, parents, and mental health professionals make informed decisions.

Additionally, this project is highly relevant due to the scalability and adaptability of machine learning solutions in addressing this widespread issue. Machine learning algorithms can analyze vast datasets to uncover patterns and triggers that may be missed through traditional research methods. This allows for a more precise understanding of how social media usage affects mental health on an individual level. With this knowledge, the project aims to create accessible and practical tools for users to regulate their social media habits. In a world where digital platforms are constantly evolving, this approach offers a dynamic solution that can be continuously updated to keep pace with emerging trends and challenges, ensuring its long-term impact and relevance.

1.5 Methodology Used

The project methodology begins by applying machine learning techniques to social media data to identify key patterns and risk factors for mental health issues like anxiety, depression, and addiction. Data will be collected from Kaggle datasets, surveys, and questionnaires on social media habits, including screen time, content consumed, and frequency of use. This rich dataset will serve as the basis for understanding how different social media behaviors impact mental health outcomes.

Data pre-processing is a crucial next step, involving cleaning the data by handling null values and duplicates, and performing feature engineering to create meaningful variables, such as average daily screen time and exposure frequency to specific content types (e.g., negative news or comparison-based content). Once the data is prepared, it will be labeled according to mental health outcomes, such as anxiety levels, to train the ML models effectively. The project will employ both supervised and unsupervised learning techniques: a Random Forest and Support Vector Machine (SVM) will classify and predict mental health outcomes based on social media usage patterns, while K-Means clustering will identify subgroups of users with similar habits and mental health profiles. Additionally, correlation analysis and sentiment analysis will be performed to uncover strong associations between specific social media behaviors and mental health, and to assess the emotional tone of textual content (e.g., posts and comments) and its impact on mental well-being.

To evaluate the models, performance metrics such as accuracy, precision, recall, and F1-score will be used, along with cross-validation to ensure generalizability. Results will be presented through an interactive Power BI dashboard that offers personalized insights into users' social media habits. Additionally, a chatbot will provide real-time, personalized advice to encourage healthier behaviors like limiting screen time and reducing exposure to harmful content.

Chapter 2: Literature Survey

2.1 Research Papers referred

1. **Title:** Associations between Screen Time and Lower Psychological Well-Being among Children and Adolescents

   Abstract: This study investigates the relationship between screen time (including social media use) and psychological well-being among children and adolescents. The data was collected from a population-based sample to assess the impact of screen time on various aspects of well-being such as self-esteem, emotional stability, and life satisfaction.

   Inference: Excessive screen time is associated with lower psychological well-being, with negative effects on emotional health and social satisfaction, particularly among children and adolescents.

2. **Title:** Deep Learning-Based Depression Detection from Social Media: Comparative Evaluation of ML and Transformer Techniques

   Abstract: The study presents a deep learning-based approach to depression detection from social media posts. The authors compared the performance of traditional machine learning models and transformer-based models (e.g., BERT) for detecting depression using user-generated text data from social media platforms.

   Inference: Transformer techniques like BERT outperform traditional machine learning models in detecting depression from social media data, highlighting the potential of deep learning in early detection and intervention

3. **Title:** The influence of blue light on sleep, performance and well-being in young adults: A systematic review

   Abstract: This systematic review explores the effects of blue light exposure (from screens and devices) on sleep quality, cognitive performance, and overall well-being in young adults. It synthesizes findings from various studies on how blue light impacts circadian rhythms and mental health.

   Inference: Blue light exposure negatively affects sleep quality and mental well-being in young adults, causing disruptions in circadian rhythms, impaired cognitive performance, and increased risk of mental health issues like anxiety and depression.

4. **Title:** Influence of Social Media on Psychological Distress Among Youth: A Case Study of Instagram

   Abstract: This systematic review explores the effects of blue light exposure (from screens and devices) on sleep quality, cognitive performance, and overall well-being in young adults.

   Inference: Blue light exposure negatively affects sleep quality and mental well-being in young adults, causing disruptions in circadian rhythms, impaired cognitive performance, and increased risk of mental health issues like anxiety and depression.

5. **Title:** Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review

   Abstract: This review discusses how sentiment analysis and artificial intelligence techniques can be used to detect depression in social media posts. The authors evaluate various AI algorithms, such as machine learning and deep learning, for analyzing user sentiment and emotions in text data.

   Inference: Sentiment analysis using AI offers promising results in detecting depression from social media data, enabling the early identification of mental health issues. Machine learning models, particularly those leveraging text sentiment, can provide accurate predictions.

6. **Title:** The Impact of Social Media on the Mental Health of Adolescents and Young Adults: A Systematic Review

   Abstract: This systematic review summarizes the effects of social media use on the mental health of adolescents and young adults. The study focuses on depression, anxiety, and loneliness, providing insights into how prolonged and excessive use of social media can exacerbate these mental health challenges.

   Inference: Social media use is linked to mental health issues such as anxiety, depression, and loneliness in adolescents and young adults, especially when usage patterns involve overconsumption, social comparison, and passive engagement.

7. **Title:** Mental Health Disparities and Influencing Factors Among Indian Youth: A Study on External Validation, Social Media, and Demographic Variables

   Abstract: The study investigates mental health disparities among Indian youth, focusing on the impact of external validation through social media and how demographic variables contribute to these disparities. It provides an in-depth analysis of social media's role in shaping mental health outcomes.

   Inference: External validation through social media plays a critical role in shaping mental health among Indian youth. Demographic factors such as gender, socio-economic status, and urbanization also contribute to mental health disparities.

8. **Title:** Exploring the Interconnection of Social Media, Mental Health, and Youth: A Bibliometric Analysis

   Abstract: This bibliometric analysis explores the relationship between social media use and mental health issues among youth. The study maps research trends and highlights key factors contributing to the mental health challenges linked to social media.

   Inference: Research shows a growing interest in the connection between social media and youth mental health, with major factors including cyberbullying, addiction to social platforms, and the social comparison effect.

9. **Title:** Machine Learning-Based Early Detection and Intervention for Mental Health Issues in Children

   Abstract: This study explores how machine learning models can be used for the early detection of mental health issues in children. The research focuses on identifying patterns in social media use, behavioral data, and other relevant factors that correlate with early signs of mental health challenges.

   Inference: Machine learning provides promising tools for the early detection of mental health issues in children, allowing for timely interventions. The approach leverages social media usage data, offering potential for widespread application in early intervention efforts.

10. **Title:** Associations between screen-based media use and brain white matter integrity in preschool-aged children

    Abstract: This study examines how screen-based media use impacts brain white matter integrity in preschool-aged children. The study uses neuroimaging techniques to assess the correlation between media use and brain development in young children.

    Inference: Excessive screen-based media use in preschool-aged children is associated with lower white matter integrity, which may affect cognitive development, language skills, and literacy, emphasizing the need to limit screen exposure at an early age.

11. **Title:** Machine Learning for Early Detection of Child Depression: A Data-Driven Approach

    Abstract: This study presents a machine learning model to detect early signs of child depression based on behavioral data and social media activity. Various supervised learning techniques are employed to identify depressive patterns among children at an early stage.

    Inference: Machine learning can effectively predict child depression, enabling early intervention and treatment by leveraging behavioral data, thus improving mental health outcomes.

12. **Title:** Machine Learning Approach for Prediction of Depression of Social Media Users from Sentiment Data Using KNN, SVM and Random Forest During and Beyond COVID-19 Period

    Abstract: The study explores the use of KNN, SVM, and Random Forest algorithms for detecting depression in social media users using sentiment data collected during and beyond the COVID-19 pandemic.

    Inference: Machine learning models, particularly Random Forest, show high accuracy in predicting depression from social media sentiment data, emphasizing the growing potential of AI in mental health monitoring during crises like the pandemic.

2.2 Existing Systems

1. Woebot

- Overview: Woebot is an AI-driven chatbot designed to provide mental health support by engaging users in brief, daily conversations. It monitors mental health symptoms, such as anxiety and depression, through the user's responses.
- Technology: Uses natural language processing (NLP) and machine learning algorithms to analyze conversations and provide cognitive behavioral therapy (CBT) techniques.
- Strengths: It offers real-time interventions and support, is easily accessible, and helps users self-manage mental health concerns.
- Limitations: The system focuses on textual data and conversational AI but may not capture the full spectrum of social media behaviors influencing mental health.

2. Crisis Text Line

- Overview: Crisis Text Line is a mental health support system that uses machine learning to assess real-time text-based conversations for individuals in distress. It prioritizes conversations based on the severity of the crisis and predicts suicide risk.
- Technology: The system uses NLP to analyze texts and machine learning models to predict mental health crises, allowing for real-time intervention.
- Strengths: Real-time monitoring and risk assessment. Scalable to handle thousands of conversations simultaneously.
- Limitations: It’s limited to users who proactively reach out and do not analyze passive social media activity, where early signs of mental health issues may be visible.

3. Mindstrong Health

- Overview: Mindstrong is a mobile-based app that uses digital biomarkers derived from smartphone activity, such as typing speed, touch patterns, and social media activity, to detect mental health conditions like depression and anxiety.
- Technology: Machine learning algorithms analyze passive data collected from smartphone interactions to detect patterns related to mental health.
- Strengths: It provides a continuous monitoring approach and can identify early signs of mental health deterioration.
- Limitations: It requires continuous user interaction with the device and primarily focuses on individual-level data rather than broader social media trends.

4. AURA (AI-based Mental Health Prediction System)

- Overview: AURA is an AI-based system designed to predict mental health issues by analyzing users' social media content, such as posts and comments, for early signs of anxiety, depression, or other psychological distress.
- Technology: It uses sentiment analysis, natural language processing, and machine learning to detect emotional tones and trends in social media behavior.
- Strengths: AURA analyzes large volumes of data in real-time and can identify subtle shifts in emotional expression that may indicate mental health risks.
- Limitations: AURA's effectiveness is contingent on the type of data it is trained on, which may lead to bias or inaccurate predictions if the dataset is not diverse.

2.3 Lacuna in the existing systems

- Small sample sizes pose a significant challenge in developing machine learning models for mental health prediction using social media data. Many studies rely on limited datasets that may not adequately capture the full diversity of mental health experiences across different populations. These smaller datasets, often collected from specific platforms or regions, can lead to overfitting, where the model performs well on the training data but fails to generalize to new, unseen data. Additionally, non-representative samples tend to skew results towards certain demographic groups, such as young, tech-savvy users, leaving out older adults or those from less connected regions.

- Privacy concerns further complicate the use of social media data for mental health analysis. Since personal data shared on social media can be highly sensitive, there is a substantial risk of violating users' privacy, particularly if they are not informed about how their data is being collected and used for mental health purposes. Many platforms do not provide clear consent mechanisms or safeguards to protect users' identities, which raises ethical concerns about the exploitation of personal information. As more mental health models are integrated into social platforms, ensuring robust privacy protections becomes crucial to avoid misuse of data, maintaining user trust, and adhering to legal and ethical standards like the General Data Protection Regulation (GDPR).

- Selection bias and scalability issues also significantly impact the effectiveness of machine learning models for mental health analysis using social media data. Social media platforms are often used predominantly by younger, more tech-savvy individuals, which means that the data collected may not accurately reflect the mental health experiences of the broader population. This demographic skew leads to selection bias, where models are trained on data that doesn't account for older populations, less digitally active groups, or individuals from different socio-economic or cultural backgrounds.

2.4 Comparison of existing systems and proposed area of work

Criteria	Existing Systems	Proposed Area of Work
Focus Area	Systems like Woebot, Crisis Text Line, and Mindstrong focus on providing mental health support and real-time intervention based on user interactions.	Your project focuses on analyzing social media activity (posts, comments, behaviors) to understand and predict mental health outcomes.
Data Source	Existing systems generally use direct user input (conversations, app interactions) or text-based sentiment from social media posts (e.g., AURA, BlueDot).	You propose to explore multifaceted social media data, such as text, multimedia, engagement patterns (likes, shares), and time spent online, for a more holistic analysis of behaviors.
Machine Learning Techniques	Existing systems use techniques like NLP, sentiment analysis, CBT-based chatbots (e.g., Woebot), and predictive analytics (e.g., REACH VET).	You propose to leverage advanced machine learning algorithms (e.g., deep learning, transformer models) to predict mental health states based on behavioral patterns.
Limitations (Privacy Concerns)	Privacy risks in systems like BlueDot and Mindstrong arise from analyzing social media data, with unclear user consent for monitoring.	Your work will emphasize privacy-first approaches by anonymizing data, ensuring user consent, and possibly using differential privacy techniques to safeguard user information.
Scalability	Some systems struggle with scalability when monitoring a large number of users in real-time, like AURA and sentiment analysis tools.	You aim to develop scalable solutions that can handle large datasets across various social media platforms, optimizing models for real-time monitoring.

Table No. 1: Comparison Table of existing and proposed systems

2.5 Focus Area

The focus area of "Impact of Social Media on Mental Health using Machine Learning," lies in exploring the intricate relationship between social media usage and its effects on mental health. As the digital landscape becomes more integral to everyday life, especially for younger demographics, the constant exposure to curated content, peer comparison, and the pressure to maintain an online persona have been linked to anxiety, depression, and other mental health issues. This project aims to delve into how specific social media activities, such as the type of content consumed, time spent online, engagement patterns, and interactions with peers, contribute to these mental health outcomes. By leveraging machine learning techniques, the project seeks to identify patterns, triggers, and risk factors that can provide insights into the mental health impacts of various social media behaviors.

The project also focuses on using advanced machine learning algorithms to analyze diverse data types from social media, such as text, multimedia, and engagement metrics, providing a comprehensive analysis of user behavior. The goal is to develop predictive models that can identify early warning signs of mental health issues like anxiety, depression, and addiction based on user activity. Unlike traditional models that rely on self-reported data or direct interventions, this project will focus on passive data collection and analysis, aiming for a scalable and efficient system that can monitor users in real time and provide actionable insights. This preventive approach seeks to not only understand but also mitigate mental health risks associated with social media, providing an innovative solution to a growing global concern.

Chapter 3: Requirements

3.1 Proposed Model

The proposed model for this project combines supervised and unsupervised machine learning techniques to analyze social media usage patterns and their impact on mental health. We aim to predict mental health outcomes such as anxiety, depression, and social media addiction based on factors like screen time, content consumption, and user interactions. The core of the model relies on Random Forest and Support Vector Machine (SVM) algorithms, which will classify users into different mental health categories based on their social media behaviors. These models are well-suited for handling complex, non-linear relationships within the dataset and can provide insights into how specific usage patterns correlate with mental health outcomes.

In addition to classification, the model will employ unsupervised learning using K-Means clustering to identify subgroups of children with similar social media habits and mental health profiles. By grouping users into clusters, we can explore patterns that may not be immediately visible in the data. For example, users with high screen time and frequent exposure to comparison-based content may form a cluster more likely to experience anxiety. This clustering approach allows for a more nuanced understanding of social media's impact on different groups of users, helping to tailor interventions more effectively.

Beyond identifying patterns, the model will incorporate correlation analysis and sentiment analysis. Correlation analysis will help detect strong associations between specific behaviors, such as exposure to negative content and high screen time, and mental health issues like increased anxiety or depression. Sentiment analysis will examine the emotional tone of social media posts and comments, assessing whether exposure to negative emotions in online interactions contributes to mental health challenges.

For the model the dataset is collected from Google forms named “Impact of Social Media on Child Mental Health (Age Group 5 to 15)” . We got around 400 responses from various sources as we have created special social media pages for that task. Those responses were given by parents of children having age from 5 to 15. But as we know, to apply any model and to ensure its accuracy we require around thousands data rows. So, we created data using the authentic and popular data creation website ‘Gretel.ai’ and generated around 3000 rows. Dataset contains 14 important labels.

1. Age (5 to 15 years)
2. Gender (Male, Female)
3. OTT Subscription (Yes, No)
4. Social Media platform(Combination of the following options: a.Facebook, b.Youtube, c.Instagram, d.Snapchat, e.Tiktok, f.Others(Twitter, Linkedin, Mojo,etc.)
5. Average screen time of child on mobile phone (Less than 1 hour, 1-3 hours, 3-5 hours, More than 5 hours)
6. Type of content watched on social media (Entertainment, Educational/Inspirational, Violent/Negative, Casual scrolling)
7. Type of Mobile Game (Action, Race, Puzzle, Others)
8. Sleep Time (Before 10pm, Between 10pm and 12am, After midnight 12am)
9. Difficult to concentrate on studies (High, Moderate, Low)
10. Outdoor sports (High, Moderate, Low)
11. Sleep Issues (High, Moderate, Low)
12. Complaint about eyes getting strained (Yes, No)
13. Depressed or worried due to constant social media use (High, Moderate, Low)
14. Replicate things watched on social media (High, Moderate, Low)

These labels are taken from questions. Questions that are included in form are :

1. What is the child's age?
2. Child's gender?
3. Which school section does the child belong to?
4. Does your child have a personal smartphone?
5. Do you have any OTT subscription like Netflix, Hotstar, Amazon Prime, etc?
6. What is the avg. screen time of the child on television?
7. What type of content does the child watch on television?
8. Which social media platforms does the child commonly use?

9. What is the avg. screen time of the child on mobile phone?
10. According to your observation, what type of content does your child generally watch on social media?
11. What type of mobile games does the child play?
12. At what time does your child generally sleep at night?
13. At what time does your child generally wake up in the morning during weekends?
14. How often does your child go outside the house to play cricket, football, etc.
15. On a scale of 1 to 3, how often does the child face issues regarding sleep?
16. While sleeping, how often does your child get up by the notification sound to check the phone in the middle of the night?
17. Does your child complain about his/her eyes getting strained?
18. Does the child wear spectacles?
19. On a scale of 1 to 3, how often does the child get restless if he hasn't used social media for a while?
20. How often does it happen that while searching for some educational content on the mobile phone, the child gets distracted and instead opens some social media application?
21. Does your child find it difficult to concentrate on studies?
22. To what extent have you noticed inappropriate behavioral changes in your child due to social media usage?

These features are critical for analyzing the impact of social media on a child's mental and physical health.

Additional features of model :

- Ethical Considerations: Ensuring the project addresses data privacy and ethical concerns related to collecting sensitive mental health and social media data from children.
- Early Warning System: Building a real-time alert system within the chatbot or web app to notify parents about concerning behavioral patterns.
- Intervention Strategies: Based on model outcomes, suggest personalized intervention strategies (e.g., limit screen time, content moderation) via the chatbot or in-app notifications.

3.2 Functional Requirements

The functional requirements for this project focus on the core features that the system must deliver. First, the model must be able to collect and process data on social media usage, including screen time, content types, user interactions, and frequency of use. The system will utilize machine learning algorithms such as Random Forest and SVM to classify and predict mental health outcomes based on these inputs. Additionally, it must perform K-Means clustering to identify subgroups of users with similar social media habits and mental health profiles.

Another key functional requirement is the integration of a chatbot, which will offer real-time, personalized advice based on the user’s social media usage patterns. The chatbot should be capable of processing user input and providing recommendations to promote healthier social media habits, such as reducing screen time, taking breaks, or avoiding negative content.

3.3 Non - Functional Requirements

The non-functional requirements focus on ensuring the system operates effectively, efficiently, and securely. First, the platform should be highly scalable to handle large volumes of data, especially if it needs to analyze the social media patterns of a wide user base. It should also perform efficiently, with machine learning models trained and optimized to deliver accurate predictions and insights in real-time without significant delays.

Security and privacy are critical non-functional requirements. The system must comply with data protection regulations such as GDPR, ensuring that users' personal and social media data are securely stored and processed. Data anonymization and encryption techniques should be applied to protect sensitive information, particularly given the focus on mental health.

3.4 Hardware & Software Requirements

- Hardware:
  Server for data processing and model training

- Software:
  HTML, CSS, StreamLit: Website development

- Dialog Flow, Rasa: Chatbot Development

- Python libraries like Sklearn, Tensorflow, Pandas, NLTK, etc.: To train, test and run the ML model and apply sentiment analysis.

3.5 Technology & Tools utilized

- Excel: To view the dataset

- Power BI, Tableau: Data visualization

- MySQL: For database management

- Jupyter Notebook, VSCode: To administer & run the code

- GitHub: To deploy the code

- Google Forms: For data collection

3.6 Constraints of working

The primary constraint in this project is the availability and quality of data. Since the model relies heavily on social media usage data, obtaining comprehensive and accurate datasets can be challenging. While some data can be sourced from platforms like Kaggle, other key insights—such as detailed patterns of user interaction, specific content consumed, and mental health outcomes—may require custom data collection through surveys and questionnaires. Gathering this data from users, particularly children, introduces ethical concerns related to privacy and consent, which must be carefully managed to ensure compliance with data protection regulations like GDPR.

Another constraint is the complexity of developing a model that can generalize well across different types of users and social media behaviors. Social media usage is highly individualistic, with varied patterns based on age, personality, environment, and other factors. Designing a model that captures these nuances while avoiding overfitting can be challenging, especially when attempting to classify mental health outcomes, which are often influenced by multiple factors beyond social media use. Additionally, computational constraints may arise due to the need for processing large datasets and training complex machine learning algorithms in real-time.

Chapter 4: Proposed Design

4.1 Block Diagram of the proposed system

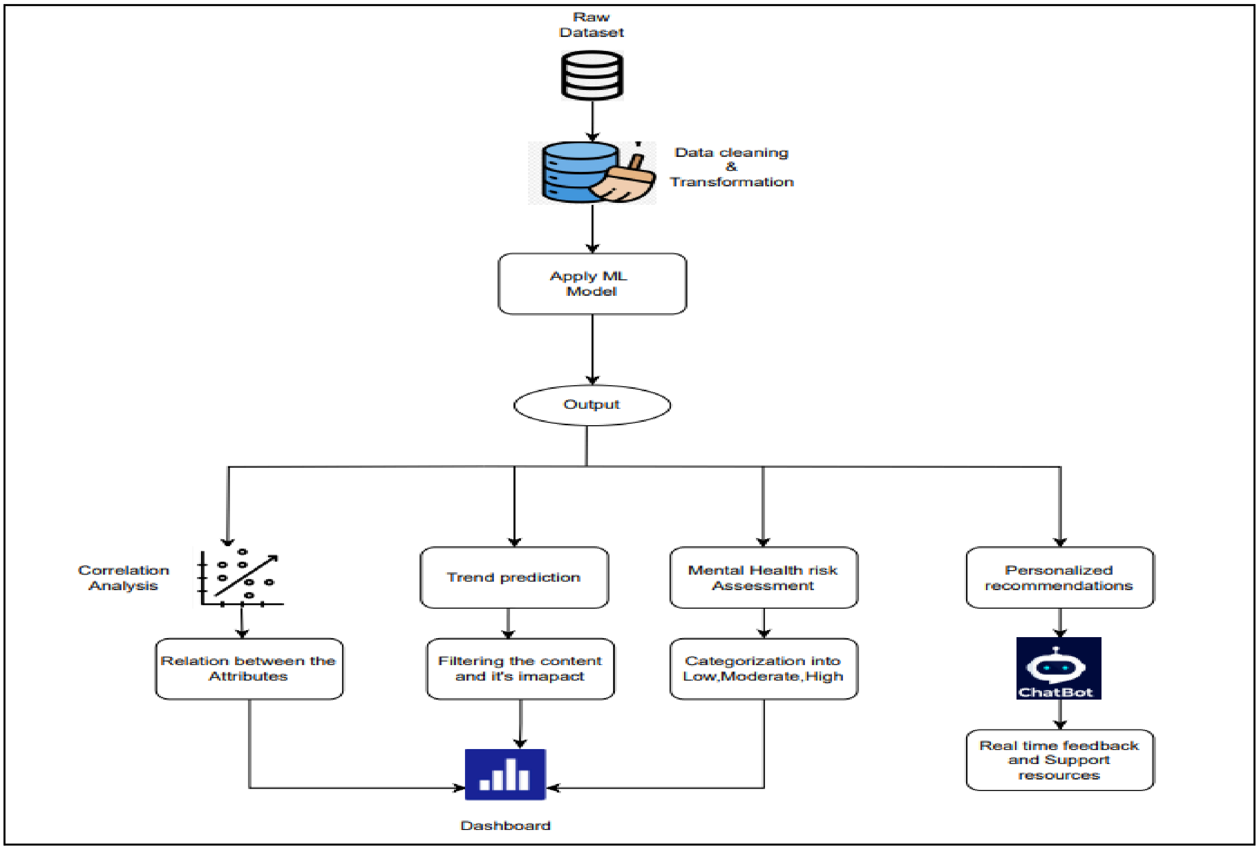

Fig.1 Block Diagram

The diagram illustrates the process of analyzing social media data to identify patterns and risk factors for mental health issues. It starts with collecting raw data from various sources, followed by cleaning and transforming the data. Machine learning models are then applied to the data to predict mental health outcomes and identify user subgroups. Correlation analysis and sentiment analysis are conducted to understand the relationships between social media behaviors and mental health.

4.2 Modular diagram

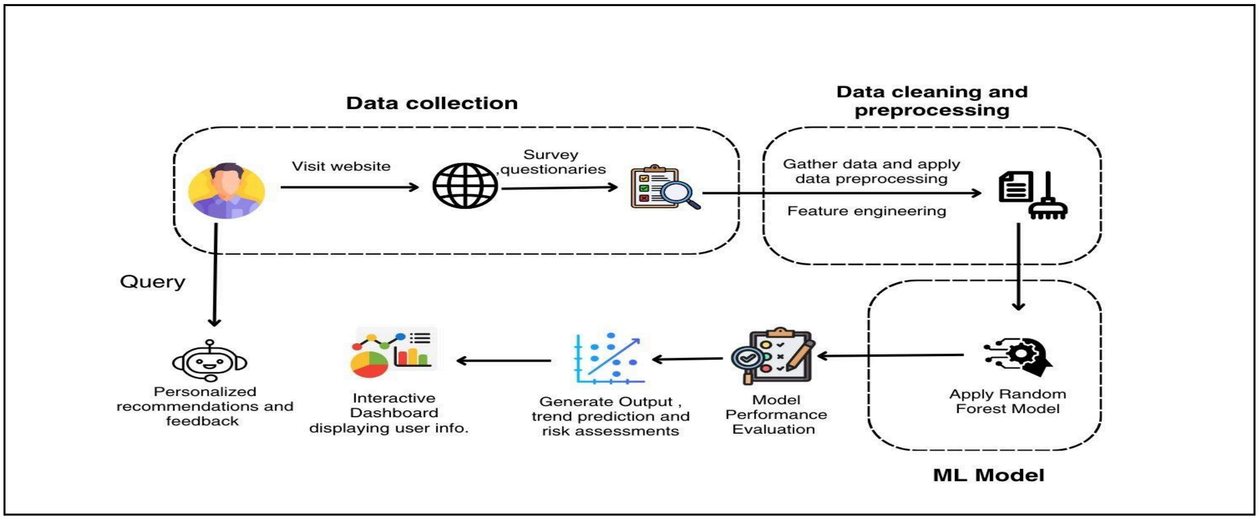

Fig. 2 Modular Diagram

In the provided diagram, the system is broken down into the following modules:

- Data Collection: This module gathers data from various sources, such as website visits and surveys.
- Data Cleaning and Preprocessing: This module cleans and prepares the data for analysis, handling tasks like removing duplicates, handling missing values, and transforming data into a suitable format.
- Feature Engineering: This module creates new features or variables from the existing data to improve the model's performance.
- ML Model (Random Forest / SVM): This module applies a Random Forest machine learning algorithm to the processed data to generate predictions or classifications.
- Model Performance Evaluation: This module assesses the accuracy and effectiveness of the ML model using appropriate metrics.
- Generate Output: This module produces the final results, such as trend predictions, risk assessments, or other relevant information.
- Interactive Dashboard: This module presents the results in a visually appealing and interactive format, allowing users to explore the data and insights.

4.3 Detailed Design Diagram

DFD level 0 diagram

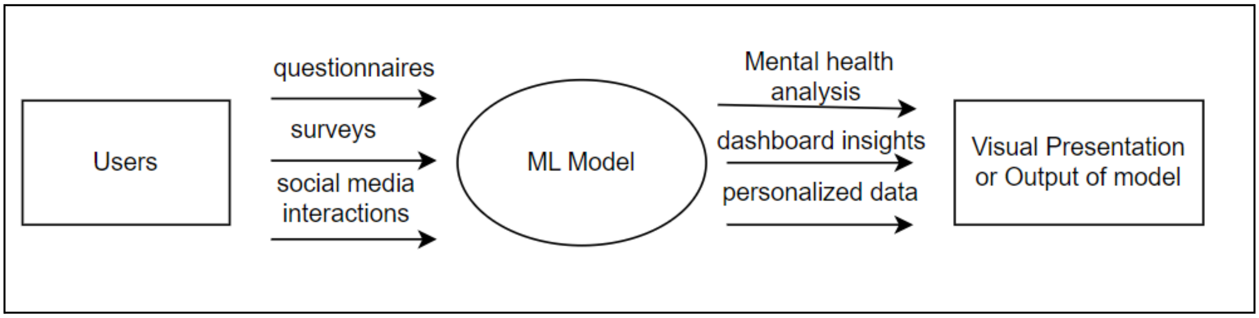

Fig. 3 DFD level 0 diagram

The DFD Level 0 diagram illustrates the high-level structure of a system for analyzing user data to provide mental health insights. Users provide data through questionnaires, surveys, and social media interactions. This data is processed by an ML model, which generates mental health analysis and dashboard insights. The results are then presented through a visual output, and personalized data is provided to the users.

In the provided diagram, we can identify the following components:

- External Entities:
    1. Users: The primary source of data for the system.
    2. Visual Presentation or Output of Model: The destination of the processed data and insights.
- Process:

  ML Model: The central component of the system responsible for analyzing data and generating insights.

State Transition diagram

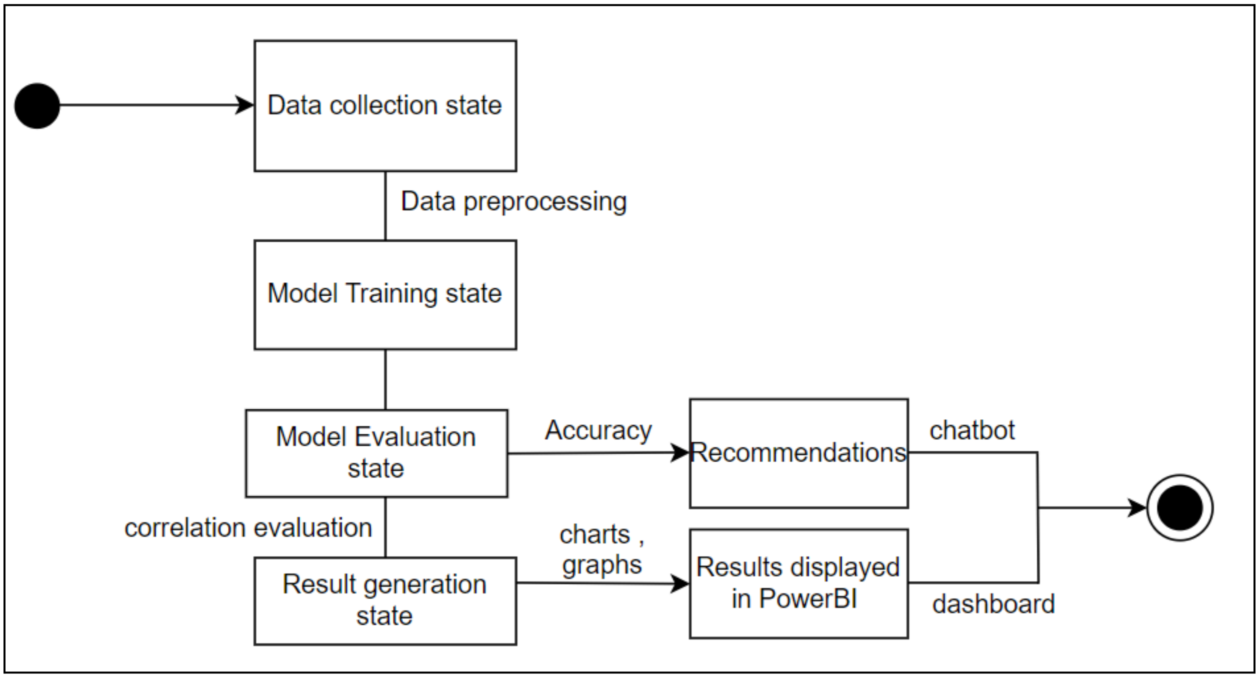

Fig. 4 State transition diagram

The state transition diagram illustrates the flow of a machine learning project from data collection to result generation and visualization. It starts with the data collection state, followed by data preprocessing, model training, and evaluation. The evaluation results lead to either recommendations or a chatbot. Finally, the results are generated and displayed in a PowerBI dashboard, along with correlation evaluation and visualization in the form of charts and graphs. This diagram provides a clear overview of the project's workflow and the transitions between different stages.

4.4. Proposed algorithms

1. Data Preprocessing Algorithm

Goal: To clean and prepare the dataset for modeling by handling missing values, duplicates, and performing feature engineering.

Steps:

- Handle Missing Data: Use imputation for numerical features (mean or median for filling null values). For categorical features, use mode or most frequent category to fill missing data.

- Feature Engineering: Calculate new features such as average daily screen time or frequency of exposure to specific content (e.g., negative content). Create new binary or categorical features based on existing attributes (e.g., categorize screen time into low, medium, and high).

.

2. Random Forest Algorithm

Goal: To classify mental health outcomes (e.g., anxiety, depression) based on social media usage patterns.

Steps:

- Initialize Parameters: Choose the number of trees (e.g., 100 trees).

- Training: At each node, the algorithm selects the best feature by calculating Gini impurity or entropy to split the data. Continue to split the data until each leaf node reaches the minimum size (or has only one class).

3. Support Vector Machine (SVM) Algorithm

Goal: To predict mental health outcomes (e.g., anxiety, depression) based on social media usage patterns.

Steps:

- Define the Kernel Function: Use a linear, polynomial, or RBF (Radial Basis Function) kernel, depending on the data distribution.

- Hyperplane Construction: Find the optimal hyperplane that maximizes the margin between the two classes (e.g., depressed vs. not depressed).

- Maximizing Margin: The support vectors (critical points close to the margin) are used to adjust the position of the hyperplane, ensuring maximum separation.

4. Correlation Analysis Algorithm

Goal: To explore relationships between social media usage patterns (e.g., screen time, type of content) and mental health outcomes (e.g., depression, anxiety).

Steps:

1. Calculate Pearson Correlation Coefficient:

- For numerical features (e.g., screen time), use Pearson correlation to measure linear relationships between variables.

2. Spearman's Rank Correlation:

- For ordinal or non-linear data, use Spearman's rank correlation to identify monotonic relationships.

3. Generate Heatmap:

- Visualize the correlation matrix using a heatmap to highlight strong positive/negative correlations.

4.5. Project Scheduling & Tracking

Date	Tasks
25/07/2024	Discussion about taking final topic
28/07/2024	Final project topic discussion
02/08/2024	Synopsis submission
17/08/2024	Literature Survey
20/08/2024	Taken guidance about PPT presentation and content adding
23/08/2024	Project Review 1
10/09/2024	Created Google form for data collection
10/09/2024 - 24/09/2024	Data collection, applying model, dashboard creation
26/09/2024	Project Review 2
27/09/2024	Data collection

Chapter 5: Results & Discussions

The results of this project will provide a detailed analysis of the relationship between social media usage patterns and mental health outcomes, particularly focusing on conditions such as anxiety, depression, and social media addiction. Using the Random Forest and SVM models, the system will classify users based on their social media habits, identifying which behaviors are most strongly correlated with negative mental health outcomes. For example, the analysis may reveal that high screen time or frequent exposure to comparison-based content significantly increases the risk of anxiety or depression. Additionally, K-Means clustering will highlight different subgroups of users who share similar social media habits and mental health profiles, offering deeper insights into how different usage patterns impact various groups. Correlation and sentiment analysis will further strengthen the understanding of how exposure to negative or emotionally charged content exacerbates mental health issues.

In discussing these results, it’s crucial to acknowledge both the benefits and limitations of the model. On the positive side, the integration of machine learning techniques has allowed for more precise and personalized predictions, helping users better understand their social media behaviors and the associated risks. The interactive dashboard and chatbot also present results in a user-friendly, actionable format, empowering users to adopt healthier digital habits. However, limitations such as the quality and availability of data, as well as potential biases in the collected data, must be considered. Furthermore, while the model offers valuable insights, mental health is influenced by a wide range of factors beyond social media, so the results should be viewed as one piece of a broader mental health picture. Overall, the findings provide a solid foundation for future research and the development of targeted interventions to improve the mental well-being of young social media users.

Chapter 6: Plan of action for the next semester

- **Chatbot:**

  Fully integrated chatbot capable of providing personalized advice based on the ML model's predictions. Research possible chatbot frameworks such as Dialog Flow, Rasa, or Microsoft Bot Framework.

- **Web Page:**

  User-friendly web interface displaying real-time insights, visualizations, and recommendations via the dashboard.Connect the backend with the ML model to serve real-time data predictions (e.g., prediction of user's mental health based on their social media usage).Integrate the chatbot into the webpage for seamless interaction.

- **Dashboard:**

  Interactive dashboard showing the correlations between social media behavior and mental health, along with personalized insights.

- **Documentation:**

  Complete documentation covering system architecture, chatbot logic, web integration, and machine learning model performance.

Chapter 7: Conclusions

The project "The Digital Mindscape: Leveraging Machine Learning to Understand Social Media's Effects on Child Mental Health" addresses the critical relationship between social media usage and mental health challenges faced by children and adolescents. By employing advanced machine learning techniques to analyze real-time social media data, the initiative seeks to uncover underlying patterns that contribute to mental health issues, such as anxiety and depression. This data-driven approach will provide personalized insights and actionable recommendations to effectively empower children and their guardians to manage social media consumption. Moreover, integrating a chatbot and an interactive dashboard will facilitate real-time engagement and support, promoting healthier social media habits. Recognizing the limitations of existing systems, this project emphasizes the need for scalability and personalization in mental health interventions. Ultimately, it strives to contribute meaningfully to the ongoing dialogue around digital well-being, providing a robust framework to mitigate the adverse effects of social media on the mental health of young users.

Chapter 8: References

Books Referred

[1] J. E. McCarthy, 2017, Social Media and Mental Health: A Comprehensive Guide, Routledge, New York, pp. 45–67.

[2] H. G. Landis, 2021, The Digital Age: Mental Health and Social Media, Academic Press, London, pp. 102–120.

Journal Papers

[3] A. R. Primack, A. J. Shensa, H. Sidani, et al., 2017, “Social Media Use and Perceived Social Isolation Among Young Adults in the U.S.,” American Journal of Preventive Medicine, vol. 53, no. 1, pp. 1–8. doi: 10.1016/j.amepre.2017.01.010.

[4] T. J. R. Naab, 2020, “The effects of social media on mental health: A systematic review,” Journal of Mental Health, vol. 29, no. 5, pp. 543–550. doi: 10.1080/09638237.2020.1748282.

[5] K. E. M. Becker, 2019, “Understanding the impact of social media on mental health: A review,” Journal of Social Media in Society, vol. 8, no. 1, pp. 130–145. [Online]. Available: https://www.j-sms.org/8-1/

[6] P. A. H. Cheng, C. T. N. Wong, and S. M. Y. Lee, 2022, “The relationship between social media use and mental health outcomes in adolescents: A systematic review,” BMC Public Health, vol. 22, no. 1, pp. 1–9. doi: 10.1186/s12889-021-12399-1.

[7] D. D. Twenge, 2017, “Have Smartphones Destroyed a Generation?” The Atlantic, vol. 320, no. 3, pp. 1–14. [Online]. Available: https://www.theatlantic.com/magazine/archive/2017/09/has-smartphone-destroyed-a-generation/534198/.

[8] A. P. H. Primack, D. H. Shensa, and H. Sidani, 2021, “Social Media Use and Mental Health Outcomes: A Review of the Literature,” Current Opinion in Psychology, vol. 39, pp. 67–71. doi: 10.1016/j.copsyc.2020.10.003.

Chapter 9: Appendix

9.1: List of Figures

9.2: List of Tables