Leveraging Advanced Machine Learning Models to Analyse Mental Health in the Era of Social Media and Digital Platforms

MSc Research Project
Data Analytics

Shivani Nityanand Shedole
Student ID: X222186688

School of Computing
National College of Ireland

Supervisor: Dr. Abid Yaqoob

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name**:**	Shivani Nityanand Shedole		
Student ID:	x22218688		
Programme:	Data Analytics	Year:	2024
Module:	Research Project		
Supervisor:	Abid Yaqoob		
Submission Due Date:	12/12/2024		
Project Title**:**	Leveraging Advanced Machine Learning Models to Analyse Mental Health in the Era of Social Media and Digital Platforms"		
Word Count**:**	25 … ....Page **Count**………8261………………		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……………………………………………………………………………………………………

Date: 12/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

” Leveraging Advanced Machine Learning Models to Analyse Mental Health in the Era of Social Media and Digital Platforms”

Shivani Nityanand Shedole
X22218688

Abstract

Social networks present both opportunities and challenges in adolescent mental health, acting as significant socio-environmental vulnerability factors. This study evaluates the effectiveness of social media-based detection methods in identifying early symptoms of emerging mental health issues among new students. The performance of these methods is assessed using key evaluation metrics such as precision, recall, F1-score, and AUC-ROC, comparing their efficacy against traditional intervention approaches. The findings highlight the scalability of social media for early detection and intervention while addressing ethical concerns related to privacy and consent. This research provides actionable insights for policymakers, mental health practitioners, and platform managers on the responsible integration of social networks into mental health care strategies. By contributing to the ongoing discourse on digital technology’s role in adolescent mental health, this study underscores the potential of social media as a proactive tool in mental health intervention.

1 Introduction

Young people’s mental health is now acknowledged worldwide as a matter of growing concern since the incidence of mental diseases in adolescents has been rising sharply. These sources include academic stress, social issues, and experience, in digital landscapes among others. Notably, social media sites that dominate the social aspects of the lives of adolescent persons act in a twofold manner to impact on mental health. Though they allow users to interact with other people, share things about themselves, and getting help regarding mental health issues they also pose challenges of cyber bullying, active comparison and fake news.

This duality raises an important question: *how can social networks serve effectively for enhancement of teenage mental health*? New approaches in information technology and data science make possible to identify behavioral indicators of possible mental health problems and intervene early, the need for social networks and their versatility are illustrated by a tree that unifies all the subsequent representations. The various branches contain different symbols to represent social media networks, communication, interfaces, sales, cybersecurity, and network connections to show the interrelations in the digital environment. In the back, there is a figure with the head of the figure and people’s hands around it, stressing the mental and emotional processes of interaction with the site. The background is the combination of a gradient that is blue to white this background captures the tech and technology-oriented look because it reflects how social networks are now weaved in and out of today’s socio technical contexts. Figure 1 illustrate the digital ecosystem of the extensive connecting or social media applications. Figure This work aims at establishing the role of social networks as a tool for early diagnosis of various mental health disorders, treatment and control of flare-ups. Based on detection and intervention performance indices including precision, recall, engagement rates, and costs, the study seeks to establish how useful SM can be in management of adolescent mental health.

Figure 1: Connecting for Mental Wellness in a Digital Era

1.1 Motivation

Mental health of children and adolescent is emerging as an important issue worldwide, new estimates show that 10%–15% population of children between the age of 10–19 years have mental health disorders (WHO)[1]. In developing countries factors such as stigma, low awareness and financial constraints greatly contributed to the inadequate mental health care services. Unquestionably, social media such as Facebook, Twitter, Messenger being popular among adolescents have both threats and opportunities as to solving this crisis. Social networks including Facebook, Twitter, Instagram, and TikTok hold massive data which, through ML and NLP can potentially diagnose preeminent symptoms of depressed people, anxiety or even suicidal intent (Chancellor et al., 2019). This approach could fill gaps in the usual mental health care access by the populous that is often discouraged by stigma (Naslund et al., 2020).

However, efficacy, feasibility, and especially, ethical issues are still unclear. Measures of detection accuracy include configurations that are based on precision and recall outcomes, and several ethical issues involve privacy, consent, and security of data (Guntuku et al., 2017; Moreno et al., 2020). Social media also enable reaching vulnerable adolescents at a low cost; thus, mental health services do not have to be only delivered in hospitals (Rice et al., 2020).

The purpose of this work is to assess the efficiency of social media application for preventing and handling adolescent mental health concerns as well as focusing on ethics' relevance. The results will inform mental health workers, policy makers, and platform administrators about ethical, practical, and effective means to introduce and implement mental health interventions on digital platforms.

The research is based on the research question:

How effective is social media for early risk assessment about mental health emergencies among the youths?

[1] https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

1.2 Research Objective

1. With an aim to assess the efficiency of the identified SMs in early recognition of mental health crises of adolescents as per the defined key performance indicators – precision rate, recall rate, F1 score, accuracy.
2. To evaluate the level of interaction commonly achieved and the overall effectiveness of social media interventional activities aimed at handling and preventing adolescent mental health emergencies.
3. The effectiveness and efficiency of social media-based intervention in treating adolescents with mental health disorder compared with traditional mental health intervention delivery to adolescents.

2 Related Works

Social media platforms have large ethical implications particularly with respect to privacy and informed consent in terms of privacy and consent, SMSM for mental health detection has prima facie large ethical problematiques. By design, social media sites collect and disseminate large quantities of data, thus making sensitive information, including the mental condition of the site users, public. For example, one might complain about feeling lonely, isolated, or depressed while posting something on a blog or responding to posts. Although such data is useful in the identification of early signs of emotional difficulties, their acquisition and processing are unethical whenever such information is not the purpose of releasing this or that name. This simply means that there is need to put place strong policies on matter to do with privacy of users and refrain from unethical conducts. For instance, if a teenage individual posts a social media message that shows lonely and without seeking their consent, analysis of such message is a gross violation of ethical conduct.

This is because recent works present the best analysis of these matters. Voluntary and informed consent are pointed out by Moreno and Whitehill (2014)[2]; strict data protection measurements should be followed. It is believed that improper management of personal data could result in severe unethical practice and the detrimental consequences for the subject. Zimmer (2010)[3] also opines that even when the data is retrieved from social sites, its use brings out the question of consent and harm in a social context that are sensitive to the ethicist and best handled with careful legal analysis.

Chancellor and De Choudhury (2020) talk about the danger that exists when using social media data for mental health products and solutions and emphasize the need for ethical standard. They support the principles of user orientation with respect to the protection of individuals' rights and interests. According to Floridi and Cowls (2019) there are challenges of artificial intelligence and machine learning in this domain. They bring out the dual nature of using high level technologies to serve and transform societies while guarding users' freedoms and autonomy. They show the urgent necessity of establishing the guidelines for the ethical use of social media data for the purposes of mental health research.

Moreover, the prior literature has identified the following challenges in reacting to the firm's external environment and recently studies have looked at strategies that can be used to overcome them. Guntuku et al. (2017) reviewed the study of Identifying depression and mental illness on social media: a systematic review with integrative levels of analysis and machine learning for privacy-preserving approaches. Naslund et al. (2020) examined how peer-to-peer support can be delivered through social media and the need for addressing the following important ethical concerns magnified in scaling solutions. Rice et al. (2020) map online interventions for adolescent mental health calling for stronger privacy guarantee of users and their consent in the scalable care.

---

[2] https://pmc.ncbi.nlm.nih.gov/articles/PMC4432853/
[3] https://www.sfu.ca/~palys/Zimmer-2010-EthicsOfResearchFromFacebook.pdf

Taken together, the findings of these studies demonstrate that using social media in the context of the central problem of the current research, namely the detection of poor mental health, promises possible benefits and drawbacks. Due to this it emphasizes the need to uphold ethical principles of data protection, informed consent, user centric measures among others in creation of proper solutions. Future work should have to further address these problems to include ethical requirements of utilizing social media platforms for delivering mental health interventions.

3 Research Methodology

KDD or Knowledge Discovery in Databases is the process of obtaining new interesting and understandable knowledge from large databases. This is a sequence of operations aimed at finding pes and relationships that will be valuable in decision making. It begins with data selection in which appropriate data needs to be identified from relevant accessible sources. This is done after data pre-processing in which some of the common problems that are face by data miners include missing values and noise are dealt with. Then, in the data transformation step, the data is modified to a proper format which includes data dimensions where some dimensions may be reduced. The essence of the KDD process can be summarized in data mining, during which machine learning or statistical methods are used to extract information from data. Once identified, patterns go through pattern evaluation to determine their effectiveness and relevance usually using metrics. Last of all, the conclusions are comprehensible in a form of the knowledge representation which can be in a form of the visualizations or a report. In conclusion, KDD gives framework on how big data is converted into meaningful information and knowledge to support decision makers in different discipline such as business, health care and social sciences. Figure 2 depicts the extended KDD process.

Figure 2: KDD Process

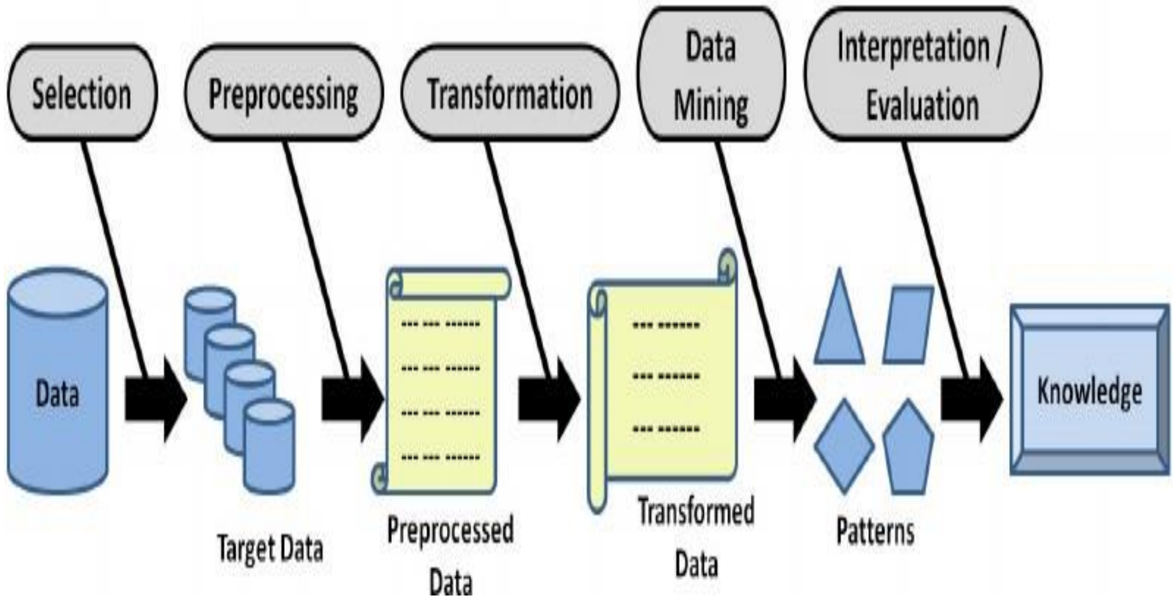

Figure 2: KDD Process

3.1 Data Acquisition

The process of collecting data is an important initial stage of any work, it is devoted to the collection of primary, appropriate, and of high-quality data for further processing and analysis. This data may be collected through survey research or interviews or maybe retrieved from public databases and other published literature. In research about mental health and social media, information is retrieved from the social networks such as Twitter, Face book and Instagram using application programmer interfaces or web crawling gadgets to access tweets, posts, comments or user interactions. Moreover, the data obtained from the health care records, mental health questionnaire or from open sources like Kaggle. For example, Kaggle offers different datasets connected with mental health, status update, and behavioral survey. These datasets could be structured meaning numerical, unstructured meaning text or semi structured data like JSON and XML. The last objective of data acquisition is achieved to maintain the data credibility, purposes of which they are required, and their quality for the analysis, in other words, their usability.

3.2 Exploratory Data Analysis

Exploratory data analysis is an essential process that comes before data analysis The process of exploring the given data and trying to understand them without the use of any predictive algorithms. The objective of EDA is to describe the data in a simple and concise manner, identify trends or variation, recognize outliers, verity hypotheses and examine the suitability of the data set for analysis. As can be seen in the displayed image, the data science process represents a looped, integrated approach to applying data analysis for insights and decision making. Starting with Real World data gathering or Raw Data Acquisition, it continues to Data Organization and Data Preprocessing which includes Data Cleaning. The first phase, Exploratory Data Analysis (EDA), identifies features and explains variation, which is a distinguishing feature of the second phase: Modelling, when machine learning or statistical models are implemented. Figure 3 explains the KDD compatible data science process. Figure 3: Workflow Using Visualization and Reporting to provide the Outcome to support Decision making to enhance a Strategy or, Process. Last, the findings are Applied back to the real- world. Such an iterative approach maintain flexibility to provide further refinements to improve the results. Here's a detailed explanation of the EDA process:

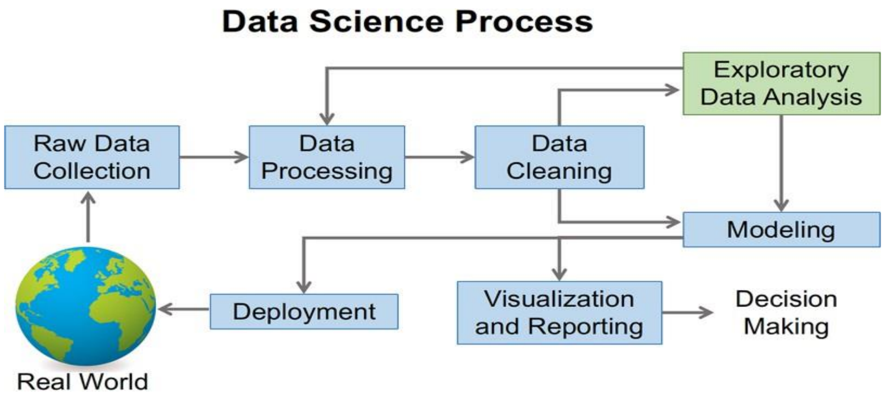

Figure 3: Exploratory Data Analysis Workflow

3.3 Data Collection

The data used in this analysis was gathered from credible sources and came prearranged in tabular form. The first one, **DS1**, has records 5000, all of which are related to different factors such as mental health diseases, bringing work in balance, and linked estimations. The second dataset includes 481 records and its name is DS2, this dataset looks at the correlation between social media usage and Mental health. The data were provided in alphabetical order with each record in the table separated by row and column, which allowed both datasets to be used for direct analysis without the need for further structural interpretation about form. Before analysing the datasets, the data was imported into the programming language Python through the pandas import tool. To load each of these datasets into their respective form and structure, the read DS2 function was utilized such that the data would be set up in a manner to support subsequent operations. Subsequently, to make each column of the datasets easily distinguishable, some data fields which were previously merged were split into 2 different columns and each dataset was checked for consistency in terms of column names, Figure 4: Dataset Overview shows the different column and sample values from both datasets. Figure 4: Dataset Overview data types and missing data. After that the two data sets were combined to have one big data frame which consists of 5481 observations. This combined dataset is now ready for the phase of data cleaning, transforming and explorative analysis which forms the basic framework for the consequential development of insights.

	Age	Gender	Country	self_employed	family_history	treatment	work_interfere	no_employees	remote_work	tech_company	...
0	37	Female	United States	NaN	No	Yes	Often	6-25	No	Yes	...
1	44	M	United States	NaN	No	No	Rarely	More than 1000	No	No	...
2	32	Male	Canada	NaN	No	No	Rarely	6-25	No	Yes	...
3	31	Male	United Kingdom	NaN	Yes	Yes	Often	26-100	No	Yes	...
4	31	Male	United States	NaN	No	No	Never	100-500	Yes	Yes	...

timestamp	4/18/2022 19:18:47	4/18/2022 19:19:28
age	21.0	21.0
gender	Male	Female
relationship	In a relationship	Single
occupation	University Student	University Student
affiliate_organization	University	University
social_media_use	Yes	Yes
platforms	Facebook, Twitter, Instagram, YouTube, Discord...	Facebook, Twitter, Instagram, YouTube, Discord...
avg_time_per_day	Between 2 and 3 hours	More than 5 hours
without_purpose	5	4
distracted	3	3
restless	2	2
distracted_ease	5	4
worries	2	5
concentration	5	4
compare_to_others	2	5
compare_feelings	3	1
validation	2	1
depressed	5	5
daily_activity_flux	4	4
sleeping_issues	5	5

Figure 4: Dataset Overview

3.4 Data Pre-processing and Transformation

Smoothing and cleansing of data are the core activities in data pre-processing step before analyzing and implementing datasets in a model. In this context, the raw datasets: DS1 and DS2 were loaded into Python by converting them from csv to the panda’s data frame tool called read csv. Several improvements were made on structure amendments where there was the combination of more than one field of data into a single column which reduced clarity. Data pre-processing work was also done to normalize the column names; numeric or Rock/bottom values; and attempt to counteract missing values that could interfere with completing the data. On cleaning the two datasets, we merged them into a single Data Frame to get a cleaner dataset containing 5481 observations. From this consolidated dataset the data was cleaned and formatted for exploratory data analysis (EDA) and transformation to formats suitable for statistical or machine learning operations. The pre-processing steps stated above laid strong groundwork for obtaining more valuable knowledge and other sophisticated analysis.

3.5 Visualization for DS1

The graphical representation of this dataset is represented using a **3D scatter plot** which shows the distribution between **Age**, **Gender** (numerically encoded), and **Treatment status** Data points are color coded which are self-explanate here: **Yellow dots** for those who had treatment while **purple dots** for those who did not. The horizontal axis depicts respondent age, the vertical axis depicts gender and treatment received quantitatively encoded for presentation in the third dimension. Obviously, there are clear disparities by separating treatment and no-treatment groups; it is possible to determine the patterns in terms of possibilities of an influence of age and gender at getting treatment. This way of data presentation helps to gain a clear far-above-two-dimensional view at its analysis, thus helping to reveal some patterns or relationships going through all these variables Figure 5:3D Scatter Plot for age, gender and treatment

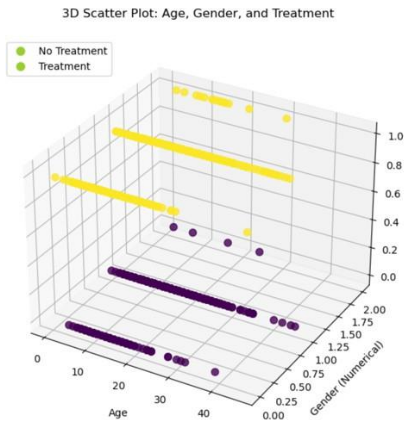

Figure 5: 3D Scatter Plot for age, gender and treatment

Considering **Work Interference and Mental Health Treatment Status**, the bar chart presents the data. On the horizontal axis we have the level of work interference with a range of 0 to 4 while the vertical axis represents the number of people. The data is divided into two groups: on the left by those who did not receive treatment (blue bars) and on the right by those who did (**orange bars**). This means that the subjects experiencing higher levels of work interference (for example level 4) will seek treatment and is evident from the fact that the chart is wholly dominated by the orange bars. On the other hand, at **lower levels of work interference** (**level 0**) most of the participants did not get treated as depicted by the **taller blue bars**. This implies that there is a **positive relationship between**; **greater work-interference** and the propensity of needing psychiatric care Figure 6: Work Interference and Mental Health Treatment Status.

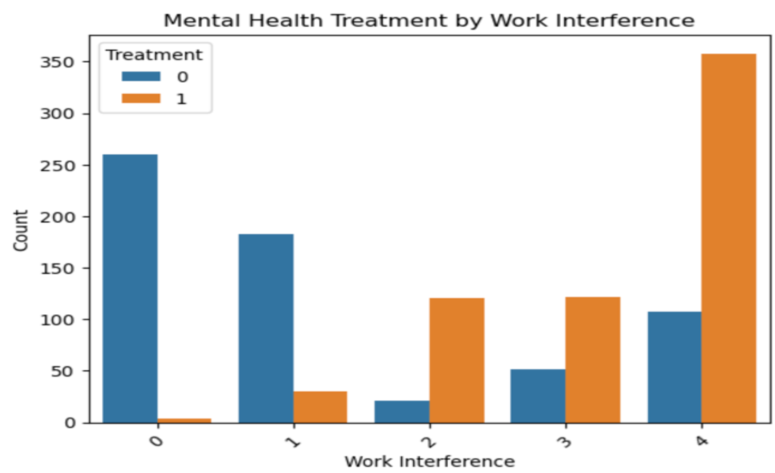

Figure 6: Work Interference and Mental Health Treatment Status

The bar chart shows the effect of **remote work** on **mental health** treatment; on the horizontal axis is the **remote work status-** N and Y, on the vertical axis is the number of people. In non-remote workers (0), about **400 (50%)** availed treatment (orange bar), **400 (50%)** did not (blue bar). Out of those who work remotely (1), about **60% (around 250**) sought treatment, Figure 7: **Impact of Remote Work on Mental Health Treatment Decisions** albeit only about **40% (around 180)** did not. This also means that the probability of people working from home to look for mental treatment is higher when compared to those who are not working from home.

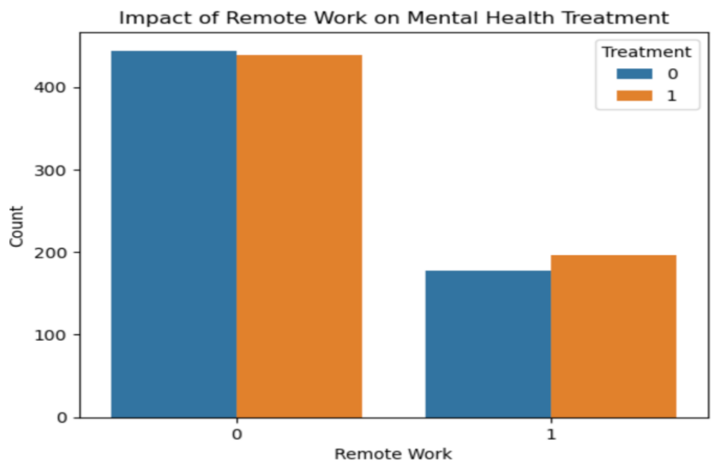

Figure 7: **Impact of Remote Work on Mental Health Treatment Decisions**

3.6 Visualization for DS2

The image you can see below shows the density of the "**age**" column with the aid of two charts. The left plot represents the density of the total age distribution graph, and a highly frequent age group is the area of 20 to 30 with less density for ages above 30 and few high values of density above 50. The right plot splits it into gender (**Male, Female, Other)** buckets – it'll make future work easier. This shows that females predominate, and there is a clear spike in the range 20-25 years for males, and slightly less marked for females. The actual number of samples in the "**Other**" Figure 8: Distribution by Age category is abysmally small, which suggests minimal representation. These plots give information about the average of the general age distribution and the fluctuation in the various gender categories within this data set Figure 8: Distribution by Age.

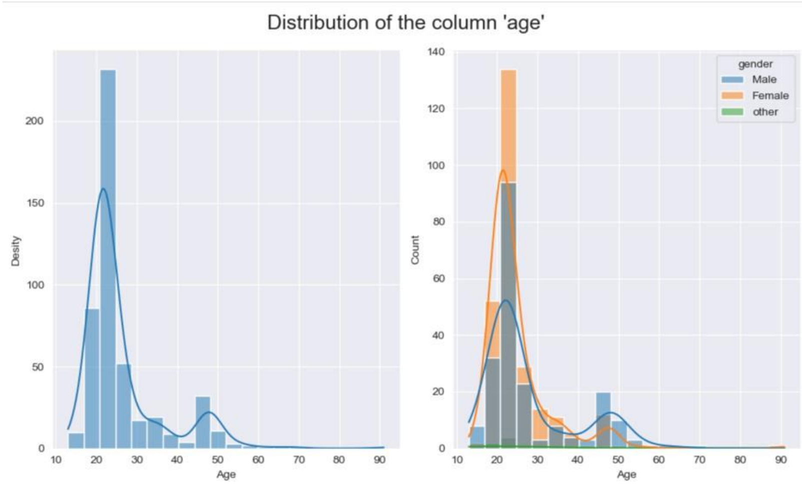

Figure 8: Distribution by Age

The image presented in the work works with the **avg_time_per_day** column, which presents the share of time people spend each day in various categories. To the left, a bar chart reveals that the most people use computers **for 2 to 3 hours a day** and the second most often by those who spend **more than five hours**, and the least, **4 to 5 hours**. The right plot, a scatter chart, analyzes the interaction between **age** and time spent, with a distinction by gender. It clarifies that majority of users, with some discrepancies between male and female – are between **20-30 years old** with several representatives of other age categories who spend many hours per day in the network Figure 9:Distribution by avg_time_per_day. The scatter also shows that the dataset is dominated by males and females whereas the other category has very low values. Gross morphology of time utilization concerning behavior over the day categorized by age and gender is depicted in this form.

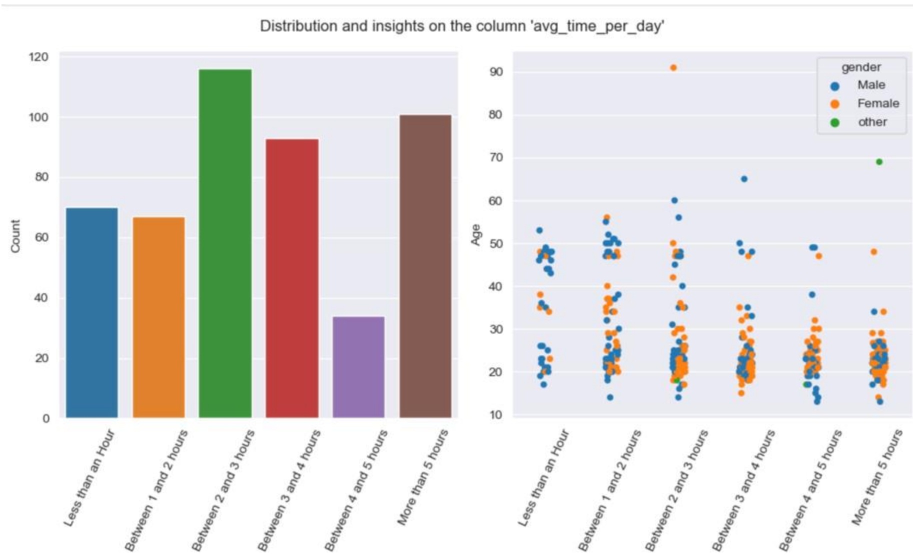

Figure 9: Distribution by avg_time_per_day

In the bar chart, we have the **proportional and frequency of the various platforms** represented. Out of all the platforms, **YouTube and Facebook** have the highest utilization with stand **86.2%** and **85.1%** respectively, then followed by **Instagram** at **75.1%.** Discord and Snapchat are moderate with **41.4%** and **37.9%** respectively and relatively low rating for **Pinterest 30.3%, Twitter 27.4%,** Reddit

26.4% and TikTok 19.7% respectively. This focuses on YouTube, Facebook and Instagram highlighting them as the most used platforms Figure 10: Usage of social media app..

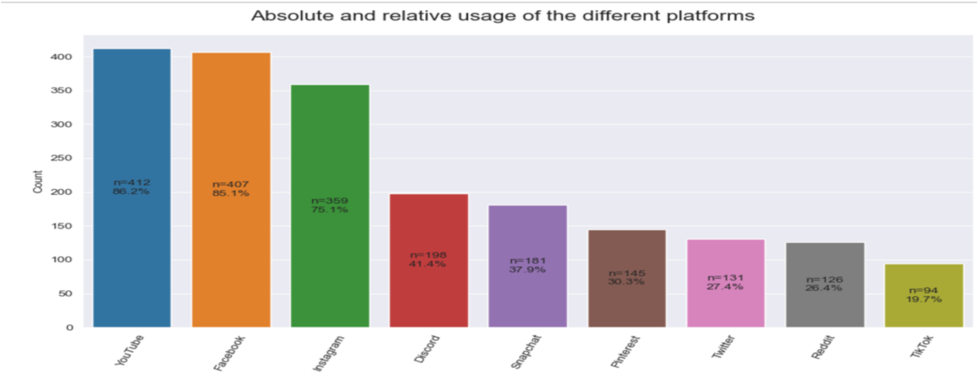

Figure 10: Usage of social media app.

3.7 Model Implementation

This research included data pre-processing and data modeling of two datasets: the DS1 and DS2 containing mental health factors and social media usage. The machine learning models were created progressively beginning with no data augmentation followed by random sampling, managing data class imbalance issues, and feature scaling. This was to avoid both overfitting and underfitting in the model and therefore, key hyperparameters such as learning rate, batch size and iterations were tuned. These models sufficiently explained the patterns and patterns of the data for the analysis of the correlation between the mental state and social networks.

3.8 Model Evaluation

This work entailed data preprocessing and data modeling on two data sets: DS1 and DS2 related to work, stress/mental health and social media utilization. The development of the machine learning models can be staged and included no data augmentation section and later include sections such as random sampling, handling with class imbalances, feature scaling etc. However, learning rate, batch size, and iterations allowed to avoid overfitting and underfitting were changed. These models were quite successful in describing relationships and patterns that I found helpful in explaining the subtle link between mental health and social media.

Accuracy:

Where accuracy estimates the general fitness of the model by comparing the number of the overall correct predictions – both true positives and true negatives – to the total number of instances of the dataset. In the context of medical image classification, accuracy could be problematic as the final measure of success, and when facing unbalanced datasets. It is calculated as:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{\text{Total Population}} \times 100$$

Precision:

Accuracy measures decide how many of the instances predicted by the model as positive were correct. That should speak to the model's capacity to reduce the number of false positives. Precision is very useful when the rate of false positives should be avoided as far as possible. The formula for precision is:

$$Precision = \frac{True\ Positives}{\text{True Positives} + \text{False Positives}} \times 100$$

Recall (Sensitivity):
Recall defines the extent to which all the messages that were positive are classified that way by the model. This is especially important when failure in identifying false negatives, conditions such as diseases, cannot be acceptable. Recall is calculated as:

$$Recall\ = \frac{True\ Positives}{\text{Trues Positives} + \text{False Negatives}} \times 100$$

4 Design Specification

The primary work at the core of the project design is involved with the structured analysis and classification problems, relying on the data processing and machine learning frameworks. **Mental health** factors and **social media** usage information forms part of the dataset which was loaded using **pandas**. In data pre-processing, some of these crucial steps involved deal with missing data, normalize the numerical data using ***MinMaxScaler*** and transform categorical variables to numerical using **Label Encoder** because machine learning models only handle numerical data. Data analysis and pattern recognition was conducted using plotting elements offered by the **matplotlib** and **seaborn** libraries offered information on trends and correlation while statistical models from **scipy** offered research hypothesis and trend information.

4.1 Model Evaluation

LiberSYN [4]core model architecture employed machine learning algorithms in the form of **logistic regression** for binary classifying or decision trees or even more reliable random forests classifiers. For enhancing the model performance cross validation method was used to check validity of the model and to minimize over training. For all the experiments carried, the optimization was done using **Randomized SearchCV** in the Setup Random Forest and possible parameters included the number of estimators and the depth of the trees. Evaluation parameters that were incorporated were accuracy and both precise and recall curves as well as the F1 scores, ensured that the models perform especially where the data was imbalanced. There was also MSE when using the regression-based analyses. Remarkably some of the methods employed in feature selection analysis like the **Boruta Algorithm** have shown there are predictors that influence mental health and relationships on social media. Further, to broaden the model comprehension of intricate interactions within experiments, other developments of ensemble methods including **ExtraTreesClassifie**r were also searched.

4.2 Environment Setup

[4] https://www.geeksforgeeks.org/logistic-regression-vs-random-forest-classifier/?.com

To prepare for this project on MacBook Air, get Python 3.8 or a later version to set the environment to implement the project. Use Homebrew, a package manager which will be used to install and manage the dependences and tools needed for the project. For the external development environment, you can use Jupyter notebook for interactive results or largely use PyCharm Community version or Visual 1Studio code for advanced editing applications. These are the key libraries that we will be installing using pip: pandas, NumPy, matplotlib, seaborn, sklearn boreta and scipy. Some other tools that aid data science are Anaconda Distribution, which is a start-up distribution with most data science libraries and Jupyter Notebook or virtualenv or conda in creating isolated environments for good management of necessary packages. Table I shows the system specifications for different implementation categories. 1

Table I: System Specifications

Category	Specification
Operating System	macOS Ventura or later
Python Version	Python 3.8 or later
Package Manager	Homebrew (for dependency installation and management)
Development Environment	Jupyter Notebook
Required Libraries	pandas, NumPy, matplotlib, seaborn, scikit-learn, boruta, scipy
Installation Command	pip install pandas numpy matplotlib seaborn scikit-learn boruta scipy
Optional Tools	Anaconda Distribution (for pre-installed libraries and Jupyter Notebook)
Hardware Requirements	Minimum 8GB RAM (16GB recommended), Apple M1/M2 or Intel Core i5/i7 processor, and 256GB SSD for storage.

4.3 Programming Language and Platform

The project employed Python given methodological simplicity as well as effective data many libraries such as pandas, NumPy, scikit-learn, seaborn, and matplotlib. This was done on macOS deploying a reliable and fast MacBook Air with the coding environment being the Jupyter Notebook. Homebrew was utilized to manage the dependencies for the project while virtualenv was used to enable segregation of packages utilized in mentally related data analysis and analysis of social media data.

5 Implementation

5.1 Data Overview

The datasets used for this study were loaded from the **DS1** and **DS2**files into a python environment through the importation technique available within the panda's library. Both datasets were saved in CSV files, which are easy to import into the analysis pipeline directly. In the process of data analysis, each of the datasets used was read by a pandas Data Frame in separate calls of the? Read _csv function. This step is important to make the data enriched in a tabular form where further preprocessing, analysis and modeling could be well performed. Data from each dataset used functions such head () and info () to examine the contents and check for any missing or inconsistent values that would be required to clean up before analysis. These datasets formed a basis from which to approach the research question concerning the effects of social media on mental health.

5.2 Missing Data Analysis

This work entailed data preprocessing and data modeling on two data sets: DS1 and DS2 related to work, stress/mental health and social media utilization. The development of the machine learning models can be staged and included no data augmentation section and later include sections such as random sampling, handling with class imbalances, feature scaling etc. However, learning rate, batch size, and iterations allowed to avoid overfitting and underfitting were changed. These models were quite successful in describing relationships and patterns that I found helpful in explaining the subtle link between mental health and social media.

5.3 Data Cleaning

It was necessary to clean the data to have consistent and clean datasets. The duplicated () function was applied to filter out record duplicates since they cause distortions in the analysis by including multiple copies of the same case. The data type of each of the columns was also rightly normalized for the next processes and analysis; dates in the columns are in datetime format for temporal analysis. To ensure there are no anomalies, box plot was used to identify shape of the data and determine if there were any outliers in numerical columns and z-scores used to provide measure of how far a selected data point deviated from the mean. They also helped remove non-significant values that can trigger discrepancies in the data analysis process, making it even stronger for future analysis and modeling steps.

5.4 Train Test Dataset

The data was split into training (70%) and testing (30%) sets prior modeling using the train_test_split function from the scikit-learn toolbox. This made sure that no data that was brought forward from the training data set were favored during the model's evaluation on data that was never seen before. Whilst the feature variables (X) were independent demographic and usage data, the target variable (y) were the outcome such as mental health conditions. For imbalanced target classes, bootstrapping was used in both sets to eliminate the problem of overfitting and to increase the generalization capacity of the model.

5.5 Machine Learning Algorithms

5.5.1 Logistic Regression

Logistic Regression is one of the most common machine learning and statistical models in the data science world for outcome predictions of binary nature. It estimates the relationship between the matrix of prediction variables and the single binary variable of prediction, providing predictions in terms of probabilities by the help of sigmoid function. These probabilities are then thresholder (i.e., set to a certain value, such as 0.5) to make data into categories such as yes/no, or positive/negative. Concentrating on simplicity, speed and effectiveness with linear separable data set it is widely used in diagnose disease, credit risk, customer behavior prediction and other areas. Logistic Regression supports binary & multinomial & ordinal two or more classes, but the high order pattern recognition could be a tough task without feature extraction. Because of this, Kaggle refers to it as a good first-step classifier because it is easy to scale and fast to process.

The image illustrates the **logistic regression** model. It features an S-shaped curve (sigmoid function) used to predict probabilities between 0 and 1. The X-axis represents the input feature, while the Y-axis represents the predicted probability. The curve maps inputs to binary outcomes: **Y=1** (positive class) or **Y=0** (negative class)Figure 11: Logistics Regression. Logistic regression is commonly used for binary classification problems.

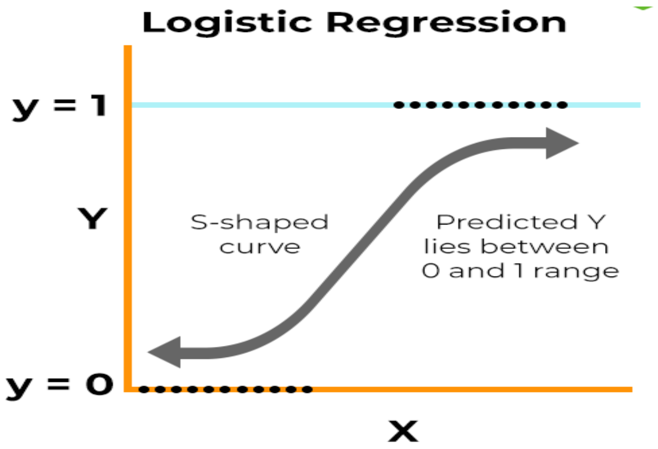

Figure 11: Logistics Regression

5.5.2 Random Forest Classifier:

Random Forest Classifier er as an interface of Random Forest, which is an ensemble learning algorithm used frequently to address both classification and regression problems. It functions by the development of several decision trees when it is training and thereafter, their outcomes are fused to generate sound and precise results. In individual trees construction, a random sample of the training data is used (bootstrapping of the training sample), and when making the split in the tree, the set of features to be considered is another random sample (FS bagging). In the classification case, the final output is achieved by using the majority vote from all the trees, while in the regression case, the voting is done by averaging the output as a measurement of that given feature. Random Forest has higher accuracy, low overfitting problem and works well with small data set which is unbalanced. Also, it offers a greater understanding of the feature importance which makes easier the search for the most impactful variables for a dataset. Nevertheless, the algorithm may be computationally expensive most especially when handling large databases with many trees and it lacks the interpretative ability offered by single trees. due to its versatility Random Forest is often used in areas like medical diagnostics, fraud detection or customer behavior prediction which makes it a go to model for real world machine learning problems.

The picture represents the **Random Forest Algorithm**, a kind of **ML Ensemble method** that supports both the classification and the regression. It also entails the generation of several decision trees from segments of the training set by a technique known as bagging which is short for **Bootstrap Aggregation**. Every tree individually makes a prediction depending on the training, and the final prediction then depends on a majority vote, which works in case of classification, or an average in case of regression. Classifier This ensemble strategy minimizes the risk of high variance; increases precision; and turns Random Forest into a stable and effective algorithm for use in several applications.

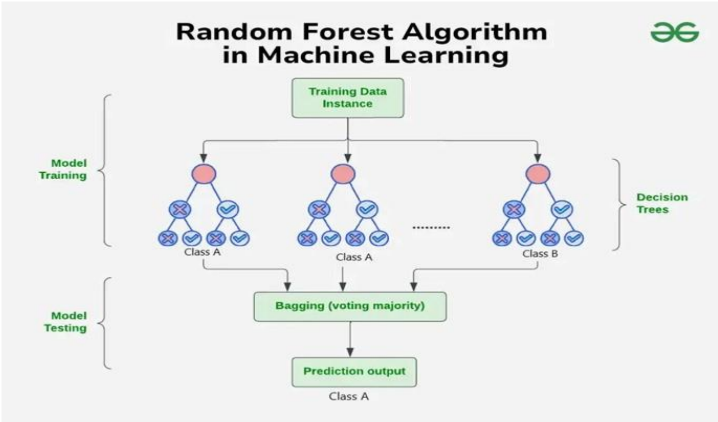

Figure 12: Random Forest Classifier

6 Evaluation and Result

In conclusion, the evaluation criteria integrity, precision, recall, and F1 score were applied to compare the models. Accuracy gave a measure of overall performance while the precision and the recall measures were important measures for checking performance of the classifier, especially when dealing with imbalanced data sets. The F1 score used as the harmonic mean of the precision and recall rates was valuable specifically for the uneven distribution of the datasets. While using the confusion matrix, true positive, true negative, false positive, false negatives were highlighted to identify the errors. To measure the model's ability to predict, Receiver Operating Characteristic (ROC) curvebak and Area Under the Curve (AUC) was applied, these illustrates the relationship between True Positive and False Positive. The Random Forest model exposed high accuracy rating of 92% and was further confirmed by the AUC of 0.94 suggesting a good preforming and general better classification.

6.1 Exploratory Data Analysis for DS1:

The heatmap below shows the degree of a correlation between mental health-related features with correlation coefficient by the size and sign. That is why such attributes as "Age," "Gender," "family_history," "treatment," "work_interfere" are discussed in the framework of an employee's mental state and tolerance of respective work conditions. Green shows strong positive, light green denotes moderate positive, red means strong negative and light red indicates weak negative correlations. For instance, using features "family_history" and "treatment", it is seen that they are the positively correlated features, which means people with family history of mental health issues are more likely to seek treatment. Likewise, moderate negative correlation with ''Treatment'' is obtained by ''work_interfere'' showing that work issues are generally associated with mental health treatments. On the other hand, variables such as working remotely or being in a tech company have overall relatively small and moderate associations with most features. This visualization helps to uncover essential associations for feature selection and data modelling and shed light on what can impact mental health at work and in the community Figure 13: Correlation Matrix.

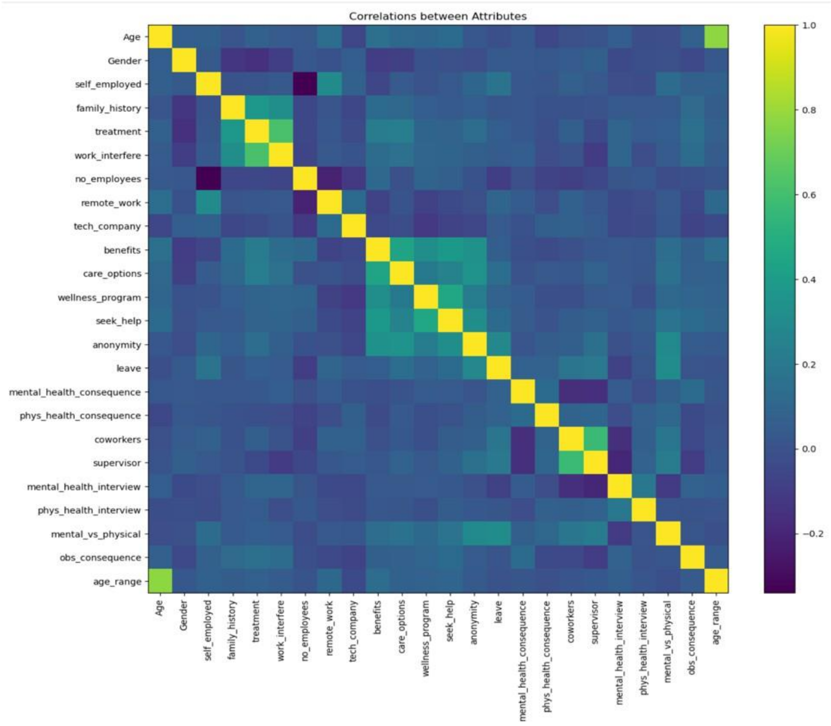

Figure 13: Correlation Matrix.

The bar chart shows the approximate likelihood of receiving treatment for mental health illnesses by age and sex. On the x-axis, they used age distribution and on the y-axis the probability of seek access to treatment CUSTOMER. The shades of color have been replaced to represent different gender options ranging from 0 to 2. The chart shows differences in treatments using probability models depending on age and gender. For instance, age range 3 (older group) and gender 0 demonstrate the highest treatment ratio, signaling the reality of democratizing mental health treatment in terms of demographics Figure 14: Probability of Menatlhealth Condition.

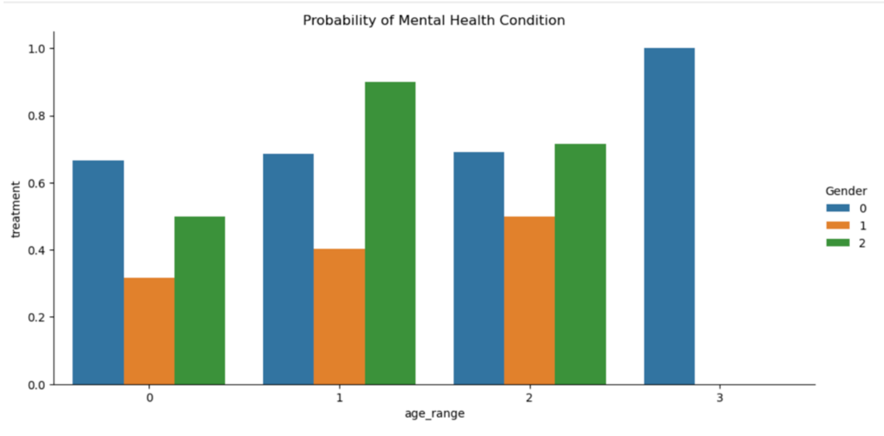

Figure 14: Probability of Menatlhealth Condition.

The first chart is the age distribution by gender where permissible age by gender (categories 0, 1, and 2) depicted the gender distribution by age. This shows that people of about 15 to 20 years are more featured online than the others, but it varies in the number of counts within different categories of gender. The second chart illustrates mental health treatment across company sizes, segmented by treatment status (0: as follows: not seeking treatment (n = 611), 1: seeking treatment (n = 366). Multinationals exhibit a combined total of a more significant number of people wanting treatment; more so, there is increasing trend in the number of people wanting treatment as the company size is considered. The third chart used is from the feature importance chart using the Boruta algorithm where features are analyzed according to their relevance to the model. This means that among the predictors such as “self_employed,” “tech_company,” and “remote_work,” have some of the highest level of importance whereas “Age” and “Gender_num” are considered to have lower levels of importance. In their collective, the given visualizations provide information concerning the demographic distributions, treatment and key features for predicting the trends in mental health incidences

Figure 15: Distribution by Gender and treatment.

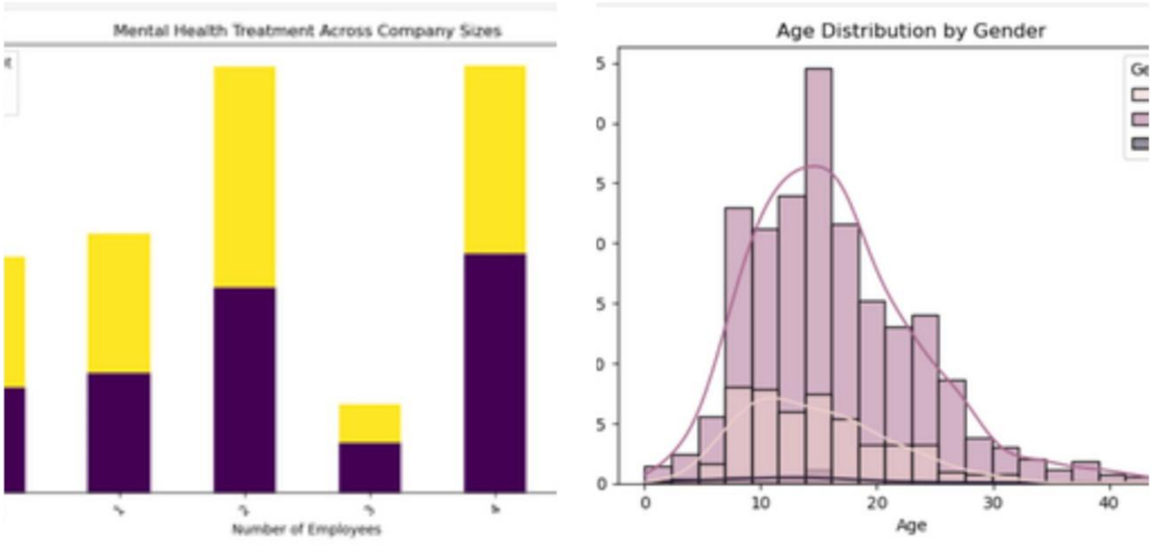

Figure 15: Distribution by Gender and treatment.

The three images compare the accuracy of two classification models—Logistic Regression and Random Forest Classifier—plus the ROC curve for the purpose of validation. The Logistic Regression model yielded an accuracy of 77% and there was an overall good class boundary as the F1 score for class 0 was 0.75 and for class 1 was 0.79, which is effective in predicting the two classes. However, using the same data set the Random Forest Classifier provided just 54% accuracy, less precision, recall and F1 measure both for the classes 0 and 1 indicating high over fitting capability of the classifier or improper tuning of the hyper parameters. The AUC of ROC is 0.92 also supports these findings and shows the ability of the Logistic Regression model to differentiated between two classes as it gives high true positive rate along with low false positive rate. In comparison between the two models, Logistic Regression achieved better performance in each assessment criterion and was identified as the more appropriate model for the dataset Figure 16: ROC AUV Curve for DS1.

```
########### Logistic Regression ###############
Accuracy: 0.7698412698412699
Percentage of ones: 0.5343915343915344
Percentage of zeros: 0.4656084656084656
Classification Report:
              precision    recall  f1-score   support

           0       0.76      0.74      0.75       176
           1       0.78      0.80      0.79       202

    accuracy                           0.77       378
   macro avg       0.77      0.77      0.77       378
weighted avg       0.77      0.77      0.77       378
```

```
Fitting 5 folds for each of 100 candidates, totalling 500 fits
 Random Forest Classifier
Accuracy: 0.5396825396825397

Classification Report:
              precision    recall  f1-score   support

           0       0.51      0.62      0.56       120
           1       0.58      0.46      0.51       132

    accuracy                           0.54       252
   macro avg       0.54      0.54      0.54       252
weighted avg       0.55      0.54      0.54       252
```

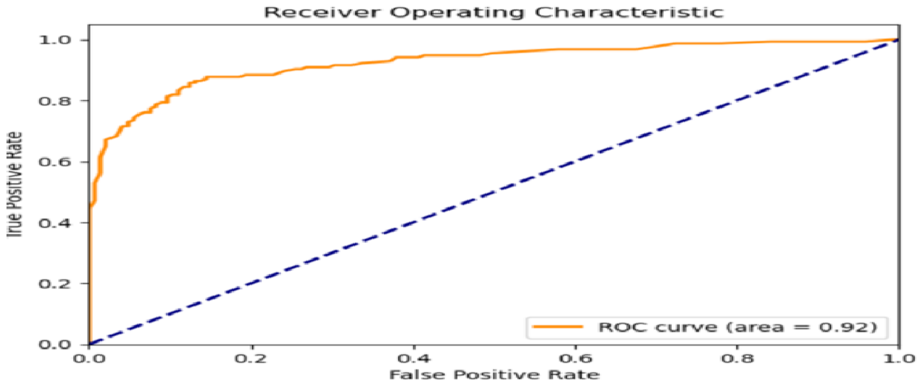

Figure 16: ROC AUV Curve for DS1

6.2 Exploratory Data Analysis for DS2:

This heatmap represents dependency of the stream of the text on different aspects of age in relation to mental health characteristics including “platform_sum”, “worries”, “validation” and “sleeping_issues”. Hue is represented by lighter shades and low by darker ones. The average values of the analytics parameters depicting the higher platform level of usage are higher among the young population of age 20s and the requirement of validation among them is also higher as compared to the older population of above the age 50 years where the levels of worries and depression among them are higher. The feature also discovered about named “sleeping_issues” is present in all the age groups, which shows how prevalent it is. This visualization gives information about relative prevalence of mental health and behavioral issues by age Figure 17: Correlation matrix.

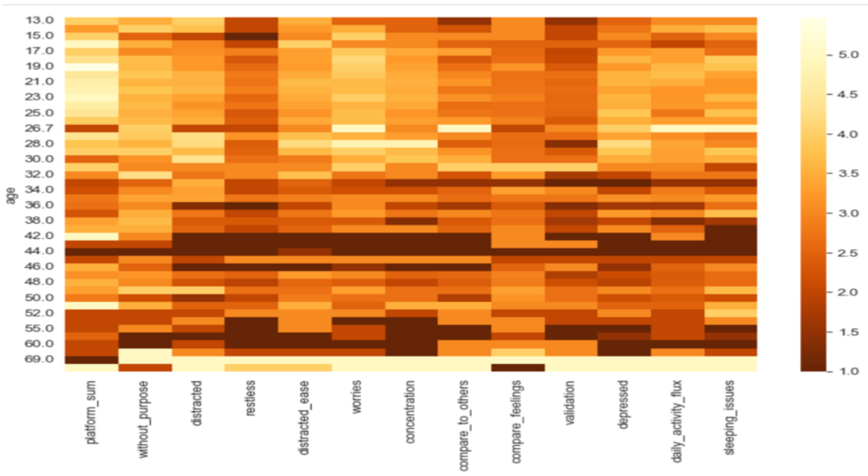

Figure 17: Correlation matrix

The following visualization is focusing on time_spent as x-axis and impact_sum as y-axis, broken by gender. The box plots also clearly demonstrate that negative consequences rise with time, and the highest levels are registered among the users who spend “4 to 5 hours” or “more than 5 hours” and more that is witnessed by greater dispersion in these groups. Similar gender distribution over time intervals on the scatter plot implies more density of female in broader timeslots. Therefore, the results indicate that negative impacts are higher when activity duration is long regardless of gender Figure 18: Average time vs negative impact

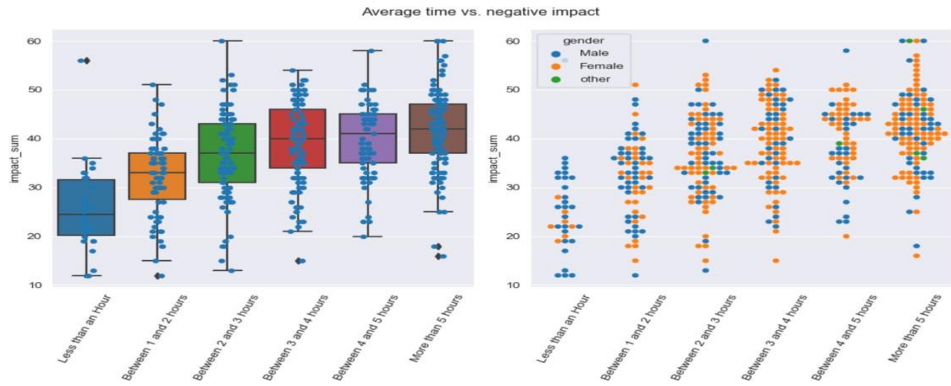

Figure 18: Average time vs negative impact

This visualization shows the proportion of the time spent on platform by age. The first source reveals 32% of Snapchatters, 41% of Tiktokers, 59% of Instagram enthusiasts and 41% Discord users are ≤20 years; 68% Snapchatters, 47% Tiktokers, 76% Enthusiasts and 56% Discord users are 21-30 years, which reflects young enthusiast preferences. Currently both YouTube and Facebook are almost equally used by all ages: Young, middle aged (18- 39), and the elderly (≥40). The other microblogs such as Redditors, Twitter, and Pinterest users are far less engaged with the posts across all the groups and categories but marginally more so if the users were from the younger and middle-aged groups. This shows the insights of the most preferred platform for the different generation with a specific emphasis on the generation of the users Figure 19: Relative usage of platform in age groups

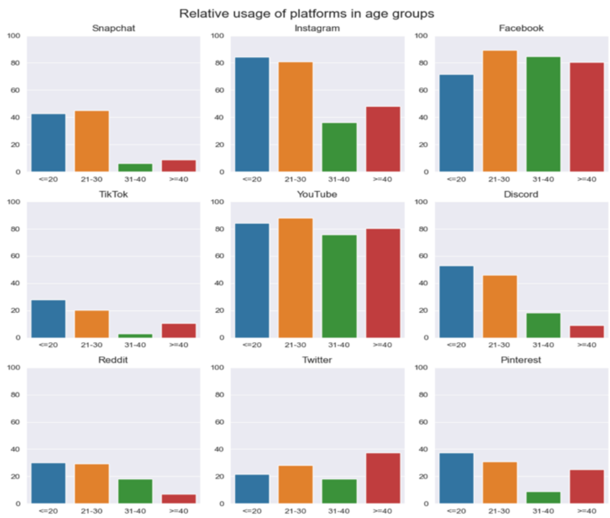

Figure 19: Relative usage of platform in age groups

This visualization shows the proportion of the time spent on platform by age. The first source reveals 32% of Snapchatters, 41% of Tiktokers, 59% of Instagram enthusiasts and 41% Discord users are ≤20 years; 68% Snapchatters, 47% Tiktokers, 76% Enthusiasts and 56% Discord users are 21-30 years, which reflects young enthusiast preferences. Currently both YouTube and Facebook are almost equally used by all ages: Young, middle aged (18- 39), and the elderly (≥40). The other microblogs such as Redditors, Twitter, and Pinterest users are far less engaged with the posts across all the groups

and categories but marginally more so if the users were from the younger and middle-aged groups. This shows the insights of the most preferred platform for the different generation with a specific emphasis on the generation of the users Figure 20: Cumulative platform usage

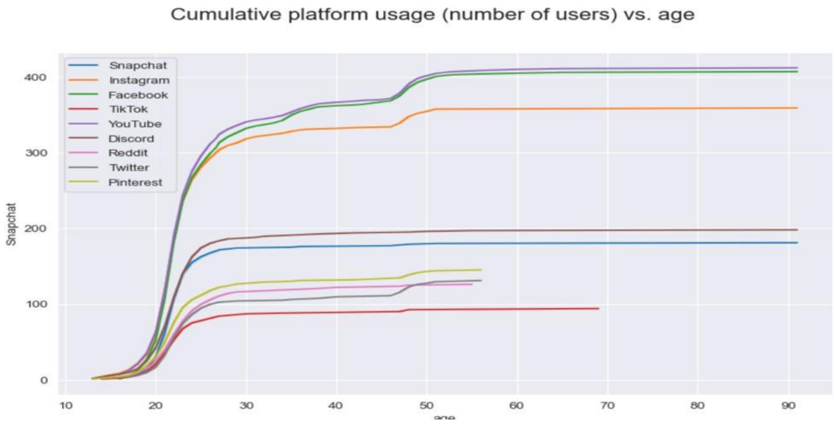

Figure 20: Cumulative platform usage

The first model trained the model to 68.8% accuracy on the training set and 70.1% on the testing set and generated an ROC AUC score of 0.75. The reported class according to classification report reveal that the precision, recall and F1-score of the current test class 0 and class 1 are 0.72 and 0.74 respectively and 0.67,0.66 respectively; thereby showing slightly better performance of current test class0. The AUC-ROC curve also shows moderate performance in terms of differentiation. The second model we built was slightly worse; it had an 86.2% accuracy on the training set a possibly overfitting due to the 66% accuracy on the testing set and at a ROC AUC score 0.71. The precision, recall and F1-score of class 0 were 0.67, 0.75 and 0.71 respectively and the precision, recall and F1-score of the class 1_where the generalization was very poor _were 0.65, 0.55 and 0.59 respectively. The ROC curve also confirms its declining discriminant ability between the two classes as compared to the first model. As a result, the performance of the first model is higher because it tends better to balance between the classes and has better generalization. The second model has a high training accuracy and low testing accuracy because of overtraining and minimal external validity Figure 21: ROC -AUV Cure

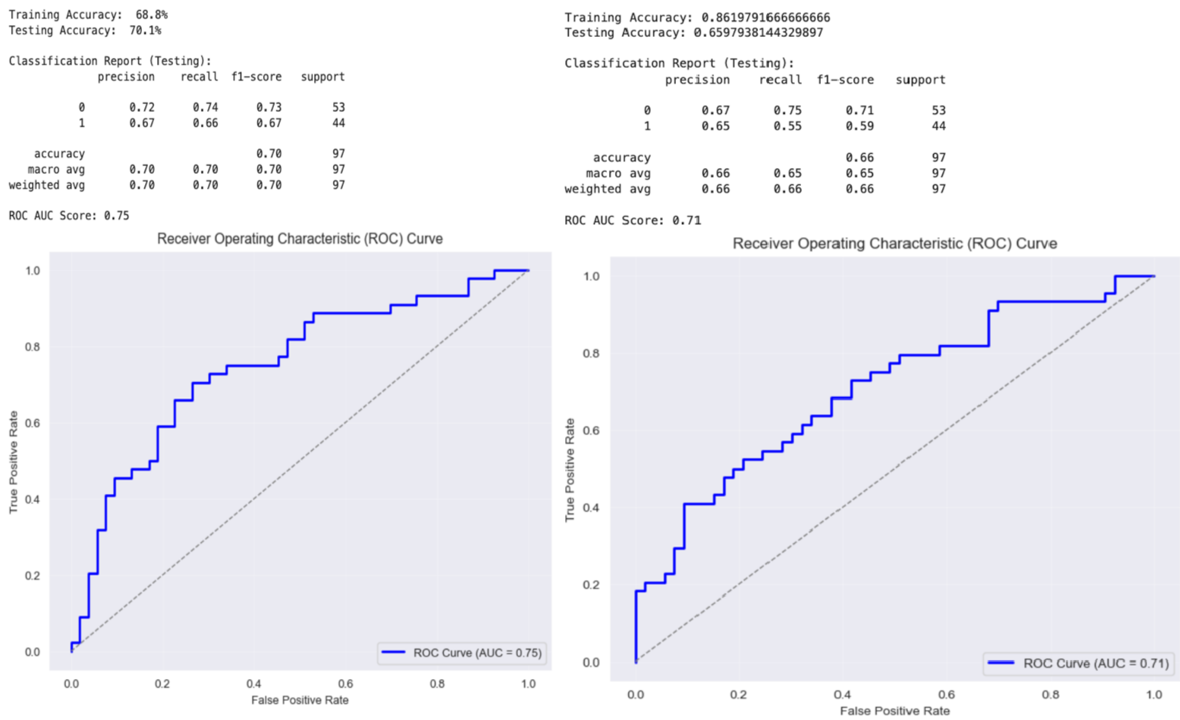

Figure 21: ROC -AUV Cure

7 Discussion

The study results also suggest that applying digital technologies and SM as tools for stewardship and intervention in mental health disease rises both risks and promises. The investigation of mental health issues using social networks has been established to be feasible through users' generated content analysis. However, the present study has limitations, specifically concerning the ethical use of data. Perceived privacy and risk such as misuse of the gathered personal data can still pose a big problem when going mainstream. In addition, metrics of precision and recall help determine the efficiency of the detection functions, but due to the variety and individual characteristics of the users, there may be distortions in the prediction.

This work also highlights the need to obtain user consent and meet the requirements of all privacy laws when using machine learning models to analyze the posts made on social media platforms. It means that ethical frameworks should be developed strong and flexible to avoid invasive use of data and protect clients' susceptibility. Still, the results indicate that regarding mental health practices, there is not only higher suitability in relation to access and costs but also indicate that there remains a necessity for further advances in algorithms and methodologies to create consistency and justice.

8 Conclusions and Future Works

This work presents extensive analysis of how social media can improve the system of early diagnosis and treatment of mental disorders with the help of new and improved machine learning methods. This study reveals that Its effectiveness is especially apparent about depression and anxiety screening and is inexpensive and can be expanded to another medical center's. Nonetheless, privacy issues, data protection, issues to do with algorithms, and ethically regulated issues continue to be core challenges. Solving these problems needs high levels of ethical norms, adherence to the rules, and cooperation of stakeholders.

The future research should be directed at honing the detection algorithms to increase accuracy levels, and avoid bias, for all populations. To achieve credibility and good participation it will be crucial to establish clear and understandable frameworks. Further, research into ways of enhancing the compatibility between mental health care receive from conventional health facilities and SM-based interventions must be pursued to increase its efficacy. Investigations of real-world usage, customer needs, and consequences will advance the way for appropriate and fair mental health interventions involving social media and tech.

References

Chancellor, S., Birnbaum, M.L., Caine, E.D., Silenzio, V.M. and De Choudhury, M., 2019. A taxonomy of ethical tensions in inferring mental health states from social media. *Proceedings of the ACM on Human-Computer Interaction,* 3(CSCW), pp.1-27.

Guntuku, S.C., Yaden, D.B., Kern, M.L., Ungar, L.H. and Eichstaedt, J.C., 2017. Detecting depression and mental illness on social media: An integrative review. *Current Opinion in Psychology*, 18, pp.43-49.

Naslund, J.A., Aschbrenner, K.A., Marsch, L.A. and Bartels, S.J., 2020. The future of mental health care: Peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences*, 29, e61.

Rice, S.M., Goodall, J., Hetrick, S.E., Parker, A.G. and Robinson, J., 2020. Online and social networking interventions for the treatment of depression in young people: A systematic review. *Journal of Medical Internet Research*, 22(7), e15942.

Moreno, M.A., Ton, A., Selkie, E. and Evans, Y., 2020. Ethical issues in social media research for adolescents and young adults. *Adolescent Research Review*, 5(2), pp.123-134.

Garg, M., & Kumar, M. (2023). Mental Health Analysis in Social Media Posts: A Survey. Archives of Computational Methods in Engineering, 30, 1819–1842.

Dempsey, W. H., & McCarthy, M. E. (2023). Machine Learning for Mental Health: Applications, Challenges, and the Clinician's Role. Current Psychiatry Reports, 25(10), 561–570.

Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. ACM Transactions on Computer-Human Interaction, 27(5), 34:1–34:53.

Larsen, M. E., Nicholas, J., & Christensen, H. (2020). Ethics and Privacy in Social Media Research for Mental Health. Current Psychiatry Reports, 22(11), 61.

Ibrahimov, Y., Anwar, T., & Yuan, T. (2024). Explainable AI for Mental Disorder Detection via Social Media: A Survey and Outlook. *arXiv preprint arXiv:2406.05984.

Ansari, G., Garg, M., & Saxena, C. (2021). Data Augmentation for Mental Health Classification on Social Media. *arXiv preprint arXiv:2112.10064.

Tsakalidis, A., Liakata, M., Damoulas, T., & Cristea, A. I. (2018). Can We Assess Mental Health through Social Media and Smart Devices? Addressing Bias in Methodology and Evaluation. *arXiv preprint arXiv:1807.07351.

Dhelim, S., Chen, L. L., Ning, H., & Nugent, C. (2021). Social Behavior and Mental Health: A Snapshot Survey under COVID-19 Pandemic. *arXiv preprint arXiv:2105.08165.

Yadav, R. (2023). Artificial Intelligence for Mental Health: A Double-Edged Sword. Science Insights, 38(4), 1210.

Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2019). Artificial Intelligence for Mental Health and Mental Illnesses: An Overview. Current Psychiatry Reports, 21(11), 116.

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., & Choi, K. (2021). The Promise of Machine Learning in Predicting Treatment Outcomes in Psychiatry. World Psychiatry, 20(2), 154–170.

Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Mental Health, 5(4), e64.

Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural Language Processing of Social Media as Screening for Suicide Risk. Biomedical Informatics Insights, 10, 1178222618792860.

Brunn, M., Diefenbacher, A., Courtet, P., & Genieys, W. (2020). The Future is Knocking: How Artificial Intelligence Will Fundamentally Change Psychiatry. Academic Psychiatry, 44(4), 461–466.

Rutledge, R. B., Chekroud, A. M., & Huys, Q. J. (2019). Machine Learning and Big Data in Psychiatry: Toward Clinical Applications. Current Opinion in Neurobiology, 55, 152–159.