DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,

SCHOOL OF ENGINEERING AND TECHNOLOGY,

SHARDA UNIVERSITY, GREATER NOIDA

AMIGO – YOUR SECOND SELF

A project submitted
in partial fulfilment of the requirements for the award of the degree of
Bachelor of Technology in Computer Science and Engineering

by

Srishti Singh (2018013720)

Prashant Singh (2018013967)

Harsh Gupta (2018014931)

Supervised by:

Prof. (Dr.) Arun Prakash Agrawal, Professor (CSE)

May 2022

CERTIFICATE

This is to certify that the report entitled "**AMIGO – YOUR SECOND SELF**" submitted by **Ms. Srishti Singh** (**2018013720**), **Mr. Prashant Singh** (**2018013967**) and **Mr. Harsh Gupta** (**2018014931**), to Sharda University, towards the fulfilment of requirements of the degree of **Bachelor of Technology** is a record of the bonafide final year project work carried out by them in the Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University. The results/findings contained in this project have not been submitted in part or full to any other University/Institute for the award of any other Degree/Diploma.

Signature of Supervisor

Name: **Dr. Arun Prakash Agrawal**

Designation: Professor (CSE)

Signature of Head of Department

Name: **Dr. Nitin Rakesh**

(Office seal)

Place:

Date:

Signature of External Examiner

Date:

ABSTRACT

Amigo, a web app that will help people in monitoring and coping-up with loneliness, anxiety, and stresses that they face in their daily lives. Faced with new realities in the post-covid world, Amigo will help people in maintaining their sanity by adapting to their new lifestyle in a tension-free way. The application will have a psychometric questionnaire to help people track their level of Depression, Anxiety, and Stress and an anonymous chat platform where they can share their feelings and talk to each other without having the fear of being judged. An AI-powered abuse and hate tracker system is also developed, that is specifically designed to identify hate-inducing items in Hinglish language, which is one of the common languages of communication in text messages across the Indian Subcontinent, in order to flag the hate-inducing posts and ban the disruptive elements from the platform. We feel that it is high time to address the ever-growing problem of mental illness, which is now fuelled by the pandemic and isolation. We believe that our effort can be useful to a lot of people around the world, and provide a base for future research and development in the hybrid field of AI and Mental Health.

Index Terms - *Depression, Loneliness, Health, Anxiety, Stress, Psychometric Analysis, Chatrooms, Machine Learning, Web Applications.*

ACKNOWLEDGEMENT

A major project is a golden opportunity for learning and self-development. We consider ourselves very lucky and honoured to have so many wonderful people lead us through the completion of this project.

First and foremost, we would like to thank **Dr. Nitin Rakesh, HoD, CSE** who gave us an opportunity to undertake this project.

My grateful thanks to **Prof. (Dr.) Arun Prakash Agrawal** for his guidance in my project work. **Dr. Arun Prakash Agrawal**, in spite of being extraordinarily busy with academics, took time out to hear, guide and keep us on the correct path. We do not know where we would have been without his help.

CSE department monitored our progress and arranged all facilities to make life easier. We choose this moment to acknowledge their contribution gratefully.

SRISHTI SINGH (2018013720)

PRASHANT SINGH (2018013967)

HARSH GUPTA (2018014931)

TABLE OF CONTENTS

	Page No.
Title Page……...……..	i
Bonafide Certificate…….……	ii
Abstract……	iii
Acknowledgement……	iv
Table of Contents……..	v-vi
List of Figures……..	vii-viii
Timeline ……..	ix
Chapter 1: INTRODUCTION……..	1
1.1 Problem Definition…….	2
1.2 Project Overview……	3
1.3 Hardware Specifications……..……..	3
1.4 Software Specifications……..	4
Chapter 2: LITERATURE SURVEY……...	5
2.1 Existing System……..	5
2.2 Proposed System……	8
2.3 Feasibility Study……	12
Chapter 3: SYSTEM ANALYSIS & DESIGN……	15
3.1 Software Requirement Specification……..	15

3.2 Flowcharts/DFDs/ERDs……..……... 19

3.3 Design and Test Steps/Criteria……..…… 22

3.4 Module training…….. 27

3.5 Website Requirements/Deliverables…….…….. 31

Chapter 4: RESULTS AND DISCUSSION……. 34

4.1 Website User Interface……... 34

4.2 Machine Learning Classifier…….. 38

References ……... 44

LIST OF FIGURES

S. No.	Figure	Page No.
3.1	Amigo System Architecture	18
3.2	Backend Connection Error Demo	18
3.3	Logical Diagram of Amigo Application	19
3.4	UML Diagram	20
3.5	Database Schema (Amigo Chat)	21
3.6	Database Schema (Amigo Quiz)	22
3.7	Incremental Model	24
3.8	Module Analysis Breakdown	27
3.9	Import Models	28
3.10	Train Test Split Distribution	28
3.11	Random Forest Classifier	29
3.12	Logistic Regression	29
3.13	Decision Tree Classifier	29
3.14	Ada Boost Classifier	30
3.15	XG Boost Classifier	30
3.16	Support Vector Machine	30

3.17	Gaussian Naive Bayes	31
3.18	Message Report Workflow	33
4.1	Sign-Up Page	34
4.2	Responsive UI – Mobile Device Screen View	35
4.3	Sign-In Page	36
4.4	Chatroom Emotional Support Channel and Threads	36
4.5	YouTube Extension Support	37
4.6	Emoji Support	37
4.7	Giphy Extension Support	38
4.8	Confusion Matrix of Random Forest Classifier	40
4.9	Confusion Matrix of Logistic Regression model	40
4.10	Confusion Matrix of Decision Tree Classifier	41
4.11	Confusion Matrix of Gaussian Naïve Bayes	41
4.12	Confusion Matrix of Support Vector Machine	42
4.13	Confusion Matrix of Ada Boost	42
4.14	Confusion Matrix of XG boost	43

TIMELINE

S.No.	Month	Brief Description	Year
1.	September	Project approved by the guide	2021
2.	October	1st evaluation done by the panel teacher and brief introduction about the project to them.	2021
3.	November	2nd Evaluation DFD, Flow Diagram, Class Diagram Presentation	2021
4.	December	Adding a feature to detect Abusive language in Chat window	2021
5.	January	Implementation of Login, Chat, Home page.	2022
6	February	Find the dataset and Data Pre-processing, Data Modelling	2022
7	March	Implementation of Support Chat Room Features and Analysis of various Machine Learning Models	2022
8	April	Working on the backend part of the project, Chat Questionnaire and Final Representation of the Project	2022

CHAPTER 1

INTRODUCTION

As the world went into lockdown because of the COVID-19 Pandemic, people were forced to give up their freedom and stay confined in their homes. The pandemic leads to the loss of loved ones, unemployment, uncertainty in financial security, homelessness and lack of social life [1]. This confinement of social animals like human beings triggered a hidden pandemic which is often neglected and looked down upon. The underlying pandemic is affecting the mental health of people from all walks of life, which is going to be a major concern in the upcoming years [2].

Alongside being a taboo, mental health is often neglected. There has been intensive research going on in the field of psychology to understand mental health and its ailments better but when it comes to combining the research with technologies like Artificial Intelligence and Machine Learning, the research is lagging behind to a great extent. The task of finding the usefulness of Artificial Intelligence and Machine Learning in the area of Mental Health is challenging and the central idea of the project.

The aim of this project is to propose a software application which can be accessible to all the people connected with the internet, and provide them with an easily accessible emotional support system and psychometric self-assessment tools to get a brief sense of their mental health anytime, anywhere, which will help them indicate to further consult a medical professional regarding their mental health if needed.

The above-stated aim can be achieved through this project by developing software that will identify the mental health software currently in use and the features they provide, and the cost associated with them. It will also look into the current research done in identifying the mental health issues using Artificial Intelligence with psychometric and demographic survey data, and further compare and use the best algorithm to make predictions for the users.

1.1. PROBLEM DEFINITION

Depression is a mood disorder that is often caused by loneliness, prolonged anxiety, or bad lifestyle behaviours. It can range from mild to severe depression which, in the worst-case scenario can lead to suicide. The current Covid-19 pandemic has only forced people to stay away from each other and avoid maintaining physical contact. This has fuelled the number of depressed patients and people are now more prone to depression and mood disorders. Oftentimes people are unaware of the symptoms of depression and fail to understand whether they have depression or not. This leads to either self-medication or ignoring the symptoms on behalf of the patients [3].

Because of the rise in the number of internet users and India and abroad, people often look towards online social communities and platforms in order to share their thoughts and avoid the feeling of loneliness. However, the dark side to these online communities and social media platforms is that they are subject to abusive and disrespectful behaviour. There can be contrasting views about the definition of abusive behaviours but these generally include personal attacks using cuss words, as well as attacking a person or an entire community based on race, gender, religion, sexual orientation or colour, etc. In the context of this paper, we define abusive language as a word or phrase, or a sentence that is meant to denigrate, offend or spread hatred towards a person or a community.

Online abusive behaviour is a societal as well as a medical problem. In a study by Munro, Emily (2011) presented that children who are subject to online abuse may develop depression, anxiety, and other mental problems. Several social media and other online platforms have added defensive measures to block offensive and toxic content from being circulated on their platform. [4] Consequently, there has been substantial research in recent years toward automated abusive language detection in the field of Natural Language Processing (NLP). The machine learning and deep learning algorithms are very much successful in providing a base for the profanity filters being used on social platforms. That said, the algorithms are not very useful for identifying abusive text-based content in languages other than English. The following problem might lead to a negative impact on the problem we are trying to solve.

1.2. PROJECT OVERVIEW

To address the problems stated above, we propose Amigo, a web app that will help people in monitoring and coping-up with loneliness, anxiety, and stresses that they face in their lives. Faced with new realities in the post-covid world, Amigo will help people in maintaining their sanity by adapting to their new lifestyle in a tension-free way. The application will have a CBT component to help people track their mental health and an anonymous chat platform where they can share their feelings and talk to each other without having to be judged. An AI-powered abuse and hate tracker system will also be installed, in order to flag the hate-inducing posts and ban the disruptive elements from the platform. We feel that it is high time to address the ever-growing problem of mental illness, which is now fuelled by the pandemic. We believe that our effort can be useful to a lot of people around the world.

1.3 HARDWARE SPECIFICATION

Table 1.1: Minimum Hardware Requirements

Operating System	Windows / Mac / Linux
Processor	Dual Core And above
RAM	2 GB RAM
Graphics Adapter	8 bits graphics adaptor and display (for 256 simultaneous colours)
Disk Space	Basic or normal disc space required by the browser like google chrome, Microsoft Edge, Opera or any other browser

1.4. SOFTWARE SPECIFICATION

Table 1.2: Minimum Software Requirements

COMPONENT	TECHNOLOGY USED
Frontend	React 17, HTML (Hypertext Markup Language), CSS, JavaScript
Backend	Node.js, Express, BCrypt
API & SDK	Stream, Twilio Messaging
Machine Learning Modules	Scikit-learn, Numpy, Pandas
Language	Python
Code Editors	Atom, Google Collab, Spyder
CI/CD Tool	Netlify
Platform	Netlify, Heroku

CHAPTER 2

LITERATURE REVIEW

2.1 EXISTING SYSTEMS

There are several existing methodologies available that work on diagnostics and help in curing depression in patients. We studied some of the existing solutions.

In the web application area, the following solutions are available in the market. Most of the applications are based on providing knowledge about mental health and the issues related to it, whereas some of them provide an option to consult a therapist but the costs of therapy are expensive. The following gives a brief review of each application under study:

A. Moodfit:

- Moodfit is a mobile application that has features such as daily goal setting and self-care methodologies.
- It also has a mood journal and a gratitude journal to keep a track of users' mood patterns.
- It also has meditation and therapy features on a paid basis.
- The app has a questionnaire that uses CBT to estimate the user's state of mindfulness.

B. Sanvello:

- Sanvello is another mobile application that uses principles of CBT.
- Provides therapy, coaching, expert help, and peer group support.
- The application is a paid application.

C. Depression CBT Self-Help Guide:

- Depression CBT Self-Help Guide provides information and tutorials to people who want to know about depression.
- It also has a feature where users can talk to an expert
- It is only available for Android devices.

D. Shine:

- Shine is another mobile application.
- Provides personalised self-care routines and a peer group for people who feel depressed.
- The app runs on a subscription model.
- Uses a questionnaire to personalise the user experience.

E. MoodMission:

- MoodMission is an application that is backed by research evidence.
- It is also a mobile application, available for Android and iOS users.
- The application uses CBT and claims to improve public mental health which is presented in research papers [5, 6].

The application of AI in detecting mental health issues has been a topic of research for quite some time and people have tried to detect mental health issues like Depression, Stress etc using various input data like tweets and threads from online social media and forums respectively [7]. Various machine learning models have been tested to predict mental health issues more precisely.

The following research has been done in the use case of Artificial Intelligence in the prediction of Mental Health Issues:

A. "Depression and Anxiety Classification of Blog Posts using CNN":

- Predicted Depression and Anxiety using Machine Learning Algorithms, in data collected from various blog posts.

- To clean and change the data from text to number applied different models like TF-IDF, and Bag-of-words.
- For all of the techniques from classification to data cleaning or pre-processing python language was used.
- CNN scores were found to be 78% (precision) and 0.72(review) [8].

B. "Prediction of Stress and Depression amongst Sea Travellers using Survey Data":

- Various ML algorithms like SVM, Lasso Regression, Random Forest Tree, And CatBoost are used for the purpose of classification.
- In this research, an interview of 470 sea travellers was done [9].
- The valuable information gathered was about the social and demographic aspects, occupations, and health of the interviewee
- It was collected through 16 labels which included academic qualifications, age, employment status, monthly salary, time of service, family structure, BMI, marital status, health records and history of the job designations and locations.
- The application of the CatBoost algorithm produced the highest accuracy at 82.6% and a precision of 84.1%.

C. "Prediction of Post-Traumatic Stress Disorder (PTSD) among Twitter Users using Hidden Markov Model":

- Reece et al. [10] worked on the prediction of Post-Traumatic Stress Disorder, commonly known as PTSD and depression among people on Twitter.
- Their Hidden Markov Model (HMM) was used to assess the presence of PTSD in an individual.
- In their dataset, 31.4% were found positive for depression and 24% for PTSD

D. "Measurement of Suicide Risks among Twitter Users using Decision Trees"

- Braithwaite et al. [11] mined twitter data of 135 participants from Amazon Mechanical Turk (MTurk)
- They applied decision tree classification to the data available to them in order to measure the suicide risks of the subjects in the dataset.

- They achieved an accuracy of 92%.

E. "Presence of Suicidal Tendencies in Tweets using CNN"

- Du et al. [12] collected Twitter data through data mining techniques and used psychiatric stressors to highlight tweets that showed suicidal tendencies.
- The CNN model outperformed the other models with a precision of 78% in identifying tweets which indicate suicidal tendencies

2.2 PROPOSED SYSTEM

2.2.1 Objectives

2.2.1.1 Overall Objectives

a. Creating a web-based, anonymous chat application where users can share their feelings with others.
b. Create a website application which will give the option of anonymous live chat to share the feelings in an emotional support room with other users without the fear of being judged.
c. Integrating a chatbot-based questionnaire to evaluate the user's mental state through a machine learning model.

2.2.2 Features of Proposed Web Application

2.2.2.1 Group Chat:

When a user registers at Amigo, he/she is provided with a unique user id and password which can be accessed on the platform. The user must verify their phone number to continue using the chat feature for safety purposes. The username for chat is system generated to ensure the anonymity of the user. If other users find the content useful, they will be rewarded with badges.

2.2.2.2 Psychometric Questionnaire:

The user may choose to talk to an AI (Artificial Intelligence) based chatbot which can also assess his mental state through his answers using the ML model. The chatbot will be able to answer most of the questions intelligently.

2.2.2.3 AI-Powered Hate and Abuse Recognition System: The hate-inducing and negative comments can be automatically detected by the system which is a model trained to detect hate speech and abusive words in the Hinglish Language. The system can then remove the hate comments and flag the users who indulged in using them. This will increase the safety of users in the chatroom.

2.2.2.4 Other Features: Other features include channel sharing positive articles, audio stories, and songs. There will also be an option to seek help from a mental health specialist with the help of direct messaging of reports to them via messaging services.

2.2.3 Methodology

2.2.2.1 Machine Learning Classifier

The proposed system evaluates a user's mental state using a Machine Learning classifier. The following shows a generic scheme of the methodology used to train a machine learning classifier.

- **Data Pre-processing**

It is the technique used to convert raw data into valuable and efficient information.

The steps involved in data pre-processing are as follows:

a. Data Cleaning:

The data can contain some missing values which can make the data set inconsistent to be used within a machine learning model. in data cleaning the missing or noisy data points are handled.

When dealing with missing data either the missing tuples are removed from the data set or are replaced to earth the mean for the most probable value.

In the case of noisy data, the outliers can be found using the following three ways binning coma regression or clustering. the outliers will fall outside the beans or clusters and then can be removed

b. Data Transformation:

The following steps can be taken to transform the data for better performance of the machine learning model it involves the following ways:

i. Attribute Selection:

The data set consists of individual scores of the psychometric test and new attributes consisting of the final score of an individual need to be calculated using the DASS-42 [13] to score evaluation for each category that is Depression, Stress and Anxiety, respectively.

ii. Normalisation:

Normalisation is done to scale the data in the range (0) to (1) or (-1) to (+ 1).

iii. Discretisation:

The numerical data can be converted to a categorical form in the final evaluated values of depression stress and anxiety using the DASS-42 [13] score for each category and then be used for classification.

- **Classification**

The term classification refers to Assigning one to be given a category to a new data point using various machine learning models. It is a supervised machine learning technique.

The following classification models were used and compared:

a. Logistic regression:

It is the most commonly used machine learning algorithm for two classifications. Easy to implement and used as the baseline for any binary classification problem.

b. Random forest classifier:

It is a widespread machine learning algorithm and it also falls into the supervised learning technique. It is chosen for both Regression and

Classification problems in machine learning. This algorithm is based on the notion of ensemble learning, which is a method of connecting numerous classifiers to crack a convoluted problem and enhance the rendition of the model.

c. Support vector machine:

It is a supervised algorithm which is mostly used for the classification or an Algorithm which is used to analyse regression and data for classification.

d. Decision tree:

It also slips underneath the type ML (Supervised learning) technique, it can be used for both classification and Regression problems and it is mainly chosen for cracking classification problems.

e. Gaussian Naive Bayes:

Gaussian Naive Bayes is a variant of Naive Bayes that follows Gaussian normal distribution and supports continuous data.

f. AdaBoost:

It is called Adaptive Boosting. It is a technique in ML seas as an Ensemble method. It used a decision tree with one level which means a decision tree with 1 split only.

g. XGBoost:

It is an ensemble learning method. Periodically, it may not stand adequate to depend upon the outcomes of just one machine learning model. Ensemble learning proposes a systematic answer to merge the predictive ability of numerous learners. The consequent is a single model which presents the aggregated outcome from numerous models.

2.2.3 Outcomes

The final proposed system will be able to do the following:

- Create a progress chart to track mental wellness scores from the questionnaire. The questionnaire results can be saved and stored for later review.
- Integrate a feature for one-on-one messaging.

- To create a safe space for everyone through the Emotional Support Chatroom.
- A safe emotional support chat room where people can anonymously share emotions without the fear of being judged or having to worry about their privacy
- Use AI-powered tracker systems to flag disruptive and abusive comments.

  With easier access to the internet few disruptive elements like trolls can abuse or induce hate in the chat room. To avoid them, an AI-powered tracker system can flag the hate inducing and abusive comments as well as there will be a provision of human moderation to catch the unwanted users who got passed the AI-powered tracking system.

2.3. FEASIBILITY STUDY

A feasibility study is defined as an evaluation or analysis of the potential impact of a proposed project. When we start a new project, we check the feasibility of the project and check whether the project is successful or not to check this we do the feasibility study

In a feasibility study we check the possible problems and check all the possible solutions to the problem and select one of the best solutions from all. In feasibility study, we collect the data from clients, their questions and their query and try to solve the problem

Any comprehension of the major specifications for the scheme is necessary for the feasibility study. In our research study, we have picked the following three points. The first is Technology. We asked questions like, Is the project technically possible? And Will failure be limited to the need for an implementation meeting the level? We have found out that it will be technically possible to build our system. Second, being Resources, We asked questions like is it financially feasible to build our system and if we have enough resources to follow on our journey. We found out we have enough resources in terms of content to refer to and finances and we have a team of three members to execute the process, this includes the human resource.

Last but not least is Time. Can we build it in a given about of time?

We planned an agile sprint model to work on given tasks to complete the project on time. Also, we need to look at what hardware and software specifications the user of this system needs

- **Technical Feasibility**

  It is concerned with specifying equipment and software that will successfully support the task required

  Is our project Technically Feasible?

  Can we make the project?

  Availability of software and technology?

  - To produce output in a given time.
  - Reaction time beneath certain conditions.
  - Capacity to process a certain transaction at a given speed.
  - Solution to communicate data to a distant location.

- **Operational Feasibility**

  It is related to human, Organization or political aspects.

  Is a human being able to use this?

  Can a human being easily operate this?

- What changes will be made in the system to easily operate
- How organizational structure will be disturbed
- What are the required new skills to easily operate?

- **Time**

  Time means time to develop the project and how much time it will take to develop the project or in that time our project is feasible or not

  Beat the competition?

  It is good to launch the project at that time?

It is a good time to build?

- **Economic Feasibility**

  Economic analysis is the most frequently used technique for evaluating the effectiveness of the proposed system. Commonly known as cost/benefit analysis

  Economic feasibility is used to check the cost of the project. It checks whether we are capable to make a profit or not.

  Is it financially practicable?

  Is it realistic for the software company and its customer or company to achieve production at a reasonable pace?

CHAPTER 3

SYSTEM ARCHITECTURE AND DESIGN

3.1 SOFTWARE REQUIREMENT SPECIFICATION

3.1.1 Product Perspective

In view of its composition, Amigo is the first of its kind and certainly the first open-source attempt to provide a platform for people to talk to others from all over the globe anonymously without being subjected to hatred or sense of being judged. This application also provides therapists support for people in need.

3.1.2 Product Functions

In Amigo app, the users can talk to each other anonymously. They can also seek the help of a therapist when they need it. The personality DASS quiz helps them access their mental health. Fun activities, music and meditation can help them improve their mood.

3.1.3 User Characteristics

3.1.3.1 Large Organizations

When fully developed, the system can be used by large scale organizations to support the mental health of their employees. Better mental health leads to improved productivity.

3.1.3.2 Academic Organizations

Students in academic organizations undergo a lot of stress for various reasons.

This application can help students to stay aware of their mental health and perform better.

3.1.3.2 Hospitals and Clinics

The hospitals and clinics can use this product to onboard their mental health patients and help them from the comfort of their homes.

3.1.4 Design and Implementation Constraints

While building and designing this application, the privacy and security of sensitive user data had to be taken care of. Hence, the application does not need permissions for a camera, location, storage, etc.

3.1.5 Assumptions and Dependencies

- There should be an internet link that is secure.
- The device should be running the latest browser.

3.1.6 Requirement Specification

3.1.6.1 User Interfaces

The user interface architecture consists of three basic interfaces, namely the signup/login window, the homepage, and the window for the DASS quiz and other activities. The components of each of the system interfaces are listed below.

The Signup/Login Window

The signup/login window offers an option for the user to create an account or log in to an existing account. There is a cache implemented to store the auth token and reduce the number of times the user needs to log in.

The Homepage

The Homepage is the page where the user can access the anonymous group chat and therapist support. There is a channels search bar and therapist search bar to list all the channels and therapists that are available.

The Quiz Window

In the quiz window, the user can access the DASS-42 Quiz and other activities.

3.1.6.2 Hardware Interfaces

The application does not need any specialized hardware interface.

3.1.6.3 Software Interfaces

The Amigo application needs a browser to run. The admin operations can be accessed in the Stream dashboard.

3.1.6.4 Communications Interfaces

As already mentioned, an Internet connection is obviously important for a complete service, but for your network adapter, you do not need any special configuration.

3.1.7 System Features

3.1.7.1 Input

The user can send messages in form of text, emojis, links (like YouTube Videos), images, videos or documents; which can be uploaded from the local device. The user can also send GIFs via using a query command and fetching the GIF results via extension.

3.1.7.2 Processing

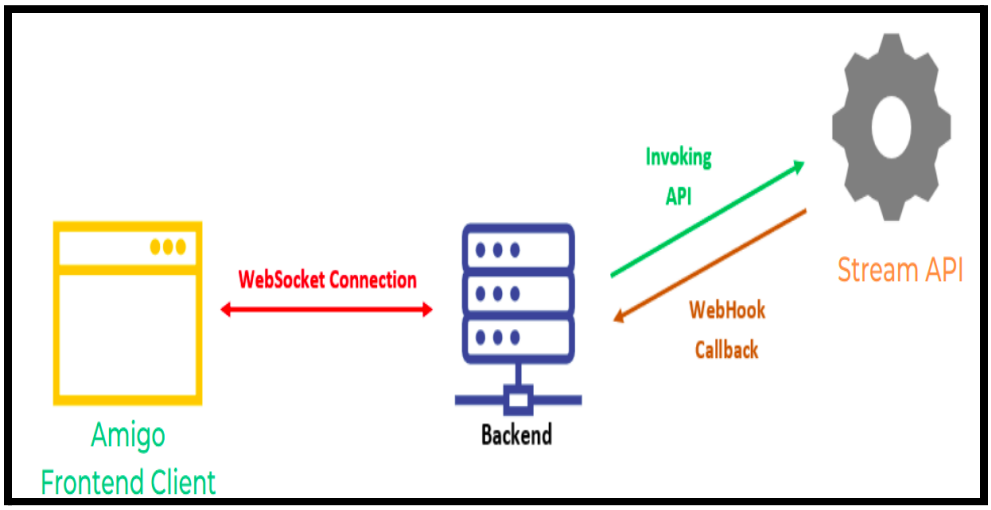

Fig 3.1: Amigo System Architecture

3.1.7.3 Error Handling

The application shows a connection error when it fails to connect to the backend of the application.

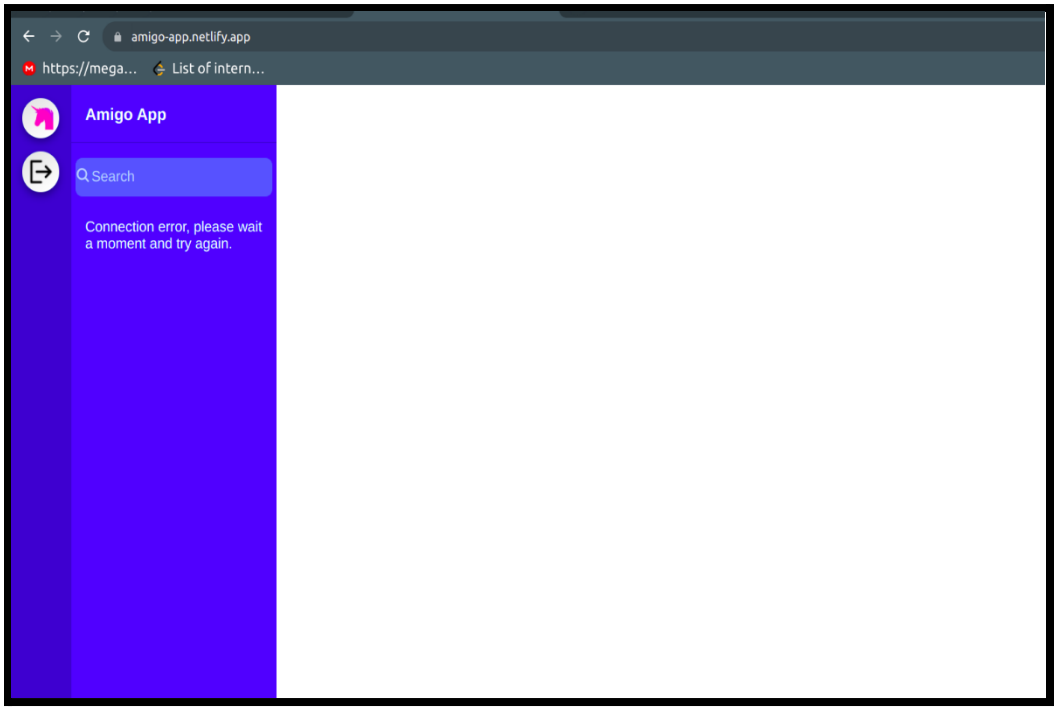

Fig 3.2: Backend Connection Error Demo

3.2 FLOWCHARTS/DFDs/ERDs

3.2.1 Data Flow Diagram:

Data flow diagram is also known as DFD. It is used to present the project process or your idea behind the project, or features of the project in graphically/Pictorial form.

It describes the process of transferring the data from the input to the file storage and reports generation.

It is divided into two categories

1. Logical

2. Physical

Data flow diagram is represented by the symbols and every symbol has a special property

- Arrow denotes the data flow.
- The circle denotes the process

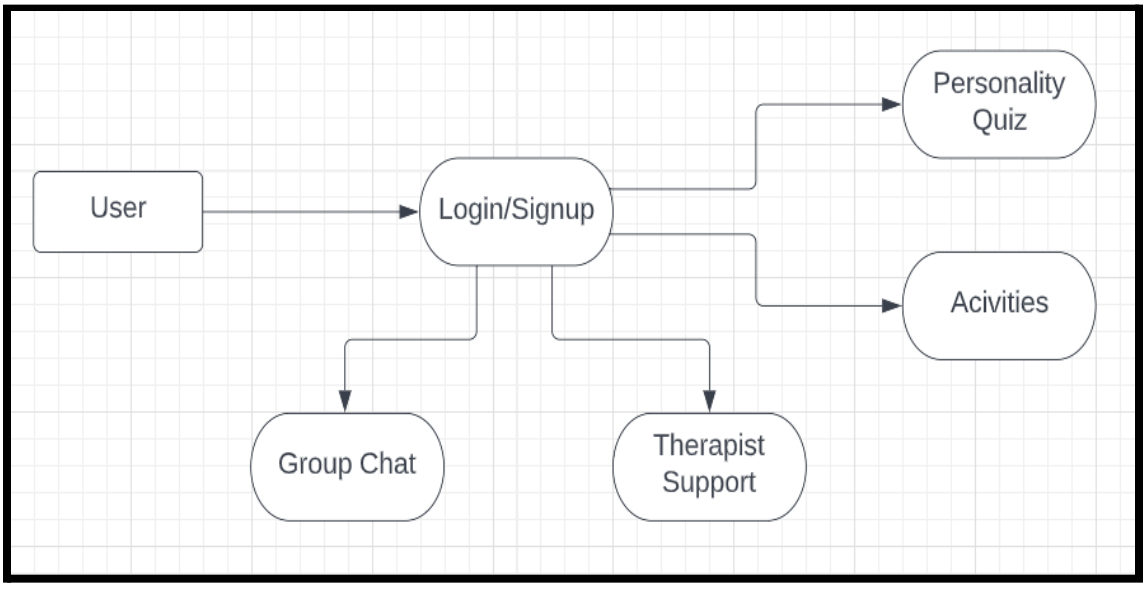

Fig 3.3: Logical Diagram of Amigo Application

3.2.2 UML Activity Diagram:

UML stands for Unified Modeling Language. The purpose of UML is to visually represent the system its roles, classes, actions and main features of the system.

We use a UML diagram to better understand the project or the working of the project to the user.

UML diagrams are divided into two categories

1. Behavioural UML Diagram

2. Structural UML Diagram

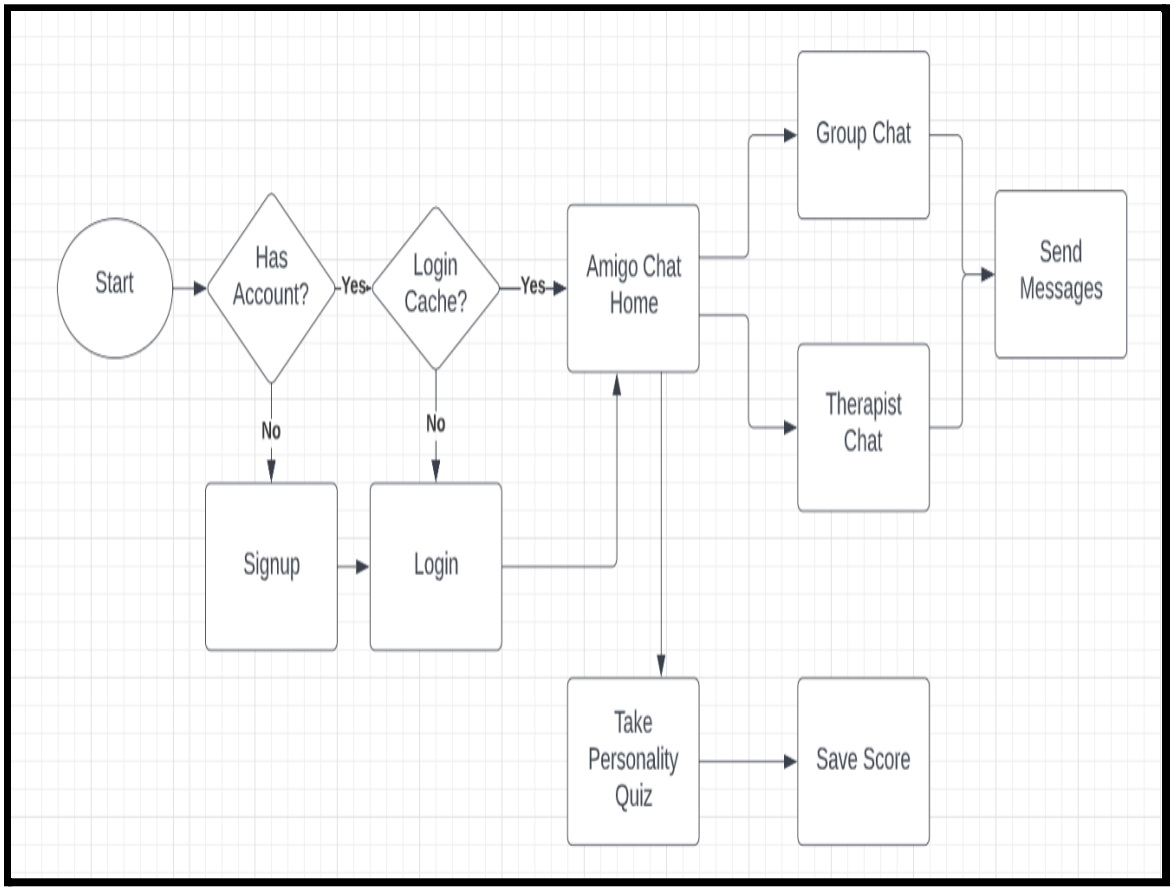

Fig 3.4: UML Diagram

3.2.3 AMIGO Database Schema

We are using a NoSQL database which is already provided in the stream SDK. The database is hosted in multiple regions to reduce latency.

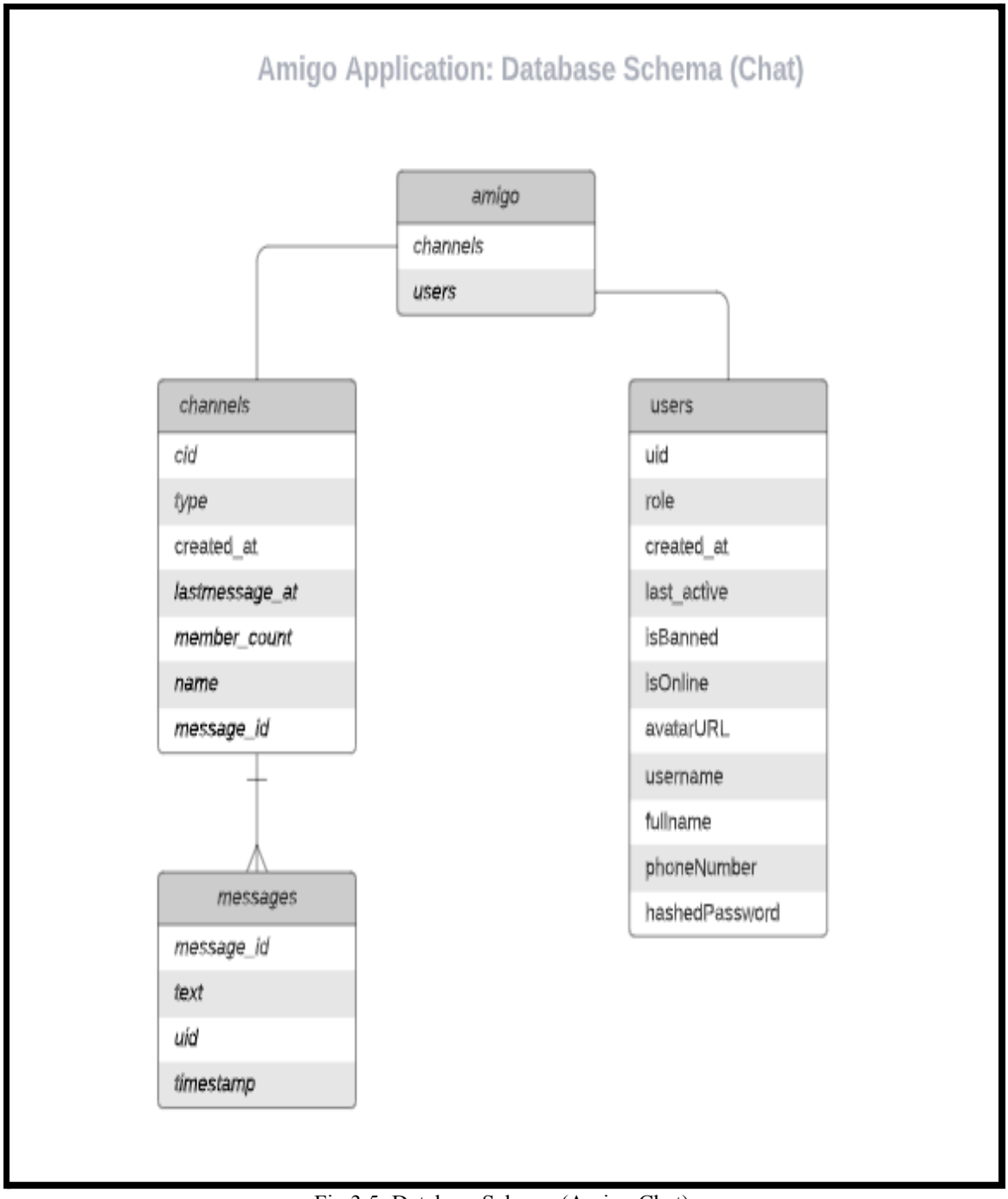

Fig 3.5: Database Schema (Amigo Chat)

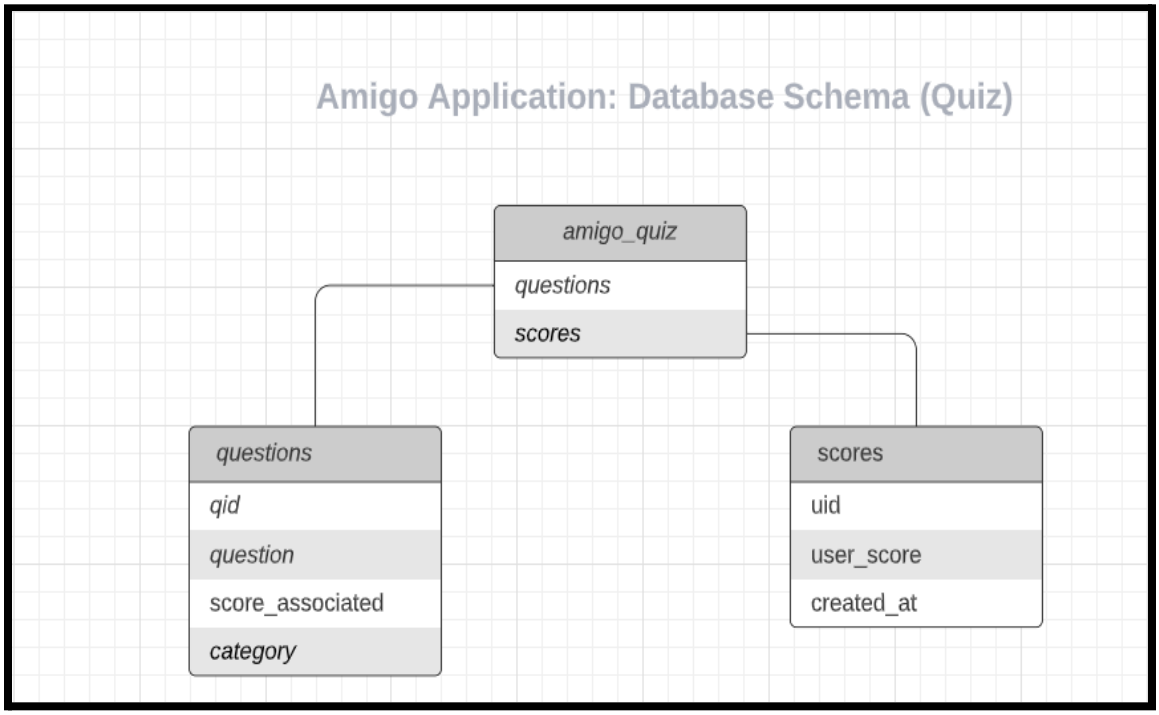

Fig 3.6: Database Schema (Amigo Quiz)

3.3 DESIGN AND TEST STEPS/CRITERIA

The following steps/phases are involved in design and planning of a project.

Phase 1: Communication

The phase of product creation begins with user and developer interactions. We also gathered the project related specifications according to work requirements.

Phase 2: System Design

A process model that is used in the implementation of the system. This activity also determines the Breakdown Structure (Modules). In the Breakdown Structure, various components used in the framework are shown.

Phase 3: Project Planning

Full calculation and timing of the entire timeline diagram for project development and for monitoring are included. Tasks are often expected to identify tools, timelines, and other details relevant to the project.

Phase 4: Modeling (Analysis & Design)

It entails a thorough review of specifications and project planning. In the analysis of demands, system analysis is done in accordance with customer requirements and what the start of the system will be in which direction it moves and what the destination will be is provided by the analysis process. In architecture, device design takes place according to research.

3.3.1 Process Model

A process is grouped into a model of the same kind by the Process Model. As a consequence, a model describes a mechanism on a type-level basis. Even though the paradigm has now reached the type stage, it is still a process of instantiation. The same method model is often used to create multiple iterations and has various instantiations. A system model should be used to prescribe how tasks can be carried out concerning the currently taking place.

The objective of a model are as follows:

• Descriptive:

1. Keep track of what occurs during a procedure.

2. Consider an outside expert's perspective who examines how an operation is carried out and determines whether changes are to be made to make it more successful or reliable.

• Prescriptive:

1. Definition of the procedures needed including how they're being performed.

2. Set laws, procedures and patterns of action that will contribute to the desired performance of the process when applied.

3. It can vary between strict adherence and fluid guidance.

• Explanatory:

1. Provide details on the rationale behind such methods.

2. Centred on logical reasoning, analysing and comparing various potential courses of action.

3. Make a strong connection between both the procedures and the standards which that model would meet.

4. Also before the positions at which tracking data can be obtained.

In this project, the incremental model was chosen.

3.3.1.1 Incremental Model

The loop model in our method is seen as an incremental solution. (Pictured) On the basis of the design and implementation of the project is chosen the S/w engineering process model is.

We have chosen an Incremental Model for our project.

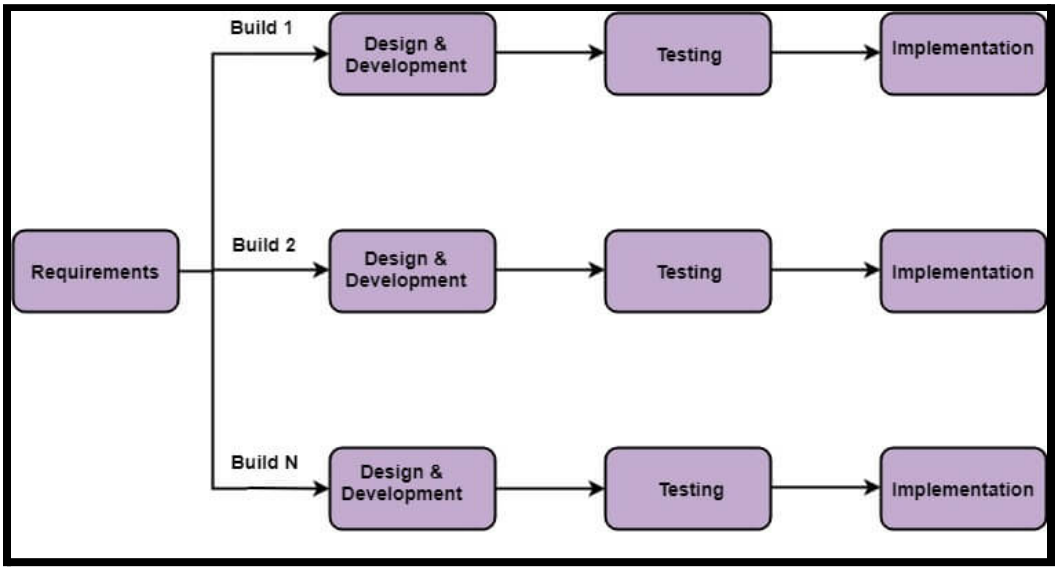

Fig 3.7: Incremental Model

A small collection of specifications is enforced easily and distributed to the authority/customer using the Incremental model.

• Changed & extended demands can be added step by step.

• It combines elements with the iterative prototyping theory of the linear sequential paradigm.

• A deliverable increment of the S/w is generated by each linear sequence.

• The Linear Sequence is divided into four sections: -

1. **Analysis:** Device & software specifications are reported and reviewed.
2. **Design:** Includes four software attributes: Data structure, S/w Architecture, representation of the interface & procedural information.
3. **Coding:** This step is used to convert the design into machine code.
4. **Testing:** Works with S/w logical internals and guarantees that all declarations are right to detect all hidden errors.

Advantages of Incremental Model:

• Generates S/w function rapidly & early during the life cycle.

• More versatile & less expensive for changing specifications.

• Easier for checking & debugging

• Customers will react to each designed product.

Why is the Incremental Approach used?

In order to boost the project's performance and usability, the key aim of using the model is to add additional features to the current modules. Using this model, we will adapt to changing consumer needs, which helps to expand the project in a very short period. The next increment in the previous raise incorporates input from consumers and several extra requirements. The process is replicated before the project is completed.

Characteristics of Incremental Model:

1. These models allow the rapid implementation and delivery of a new set of industry requirements to clients and then updating and expanding functionality step by step.
2. Each increment generates the commodity sent to the consumer and proposes certain adjustments and increments that differ with certain extra criteria compared to previous ones.
3. The radical model prevents the initiative from being completed all at once. This is useful for designing and checking components, enabling the project to be modularized for easier management.

Ultimately, the growth of the project in increments is easier. We will create a working prototype form-1 with only core tasks and then in subsequent increments, expand on this layout. Splitting the entire system into separate priority groups will serve to reduce system complexity.

3.3.2 Breakdown Structure (Modules Analysis)

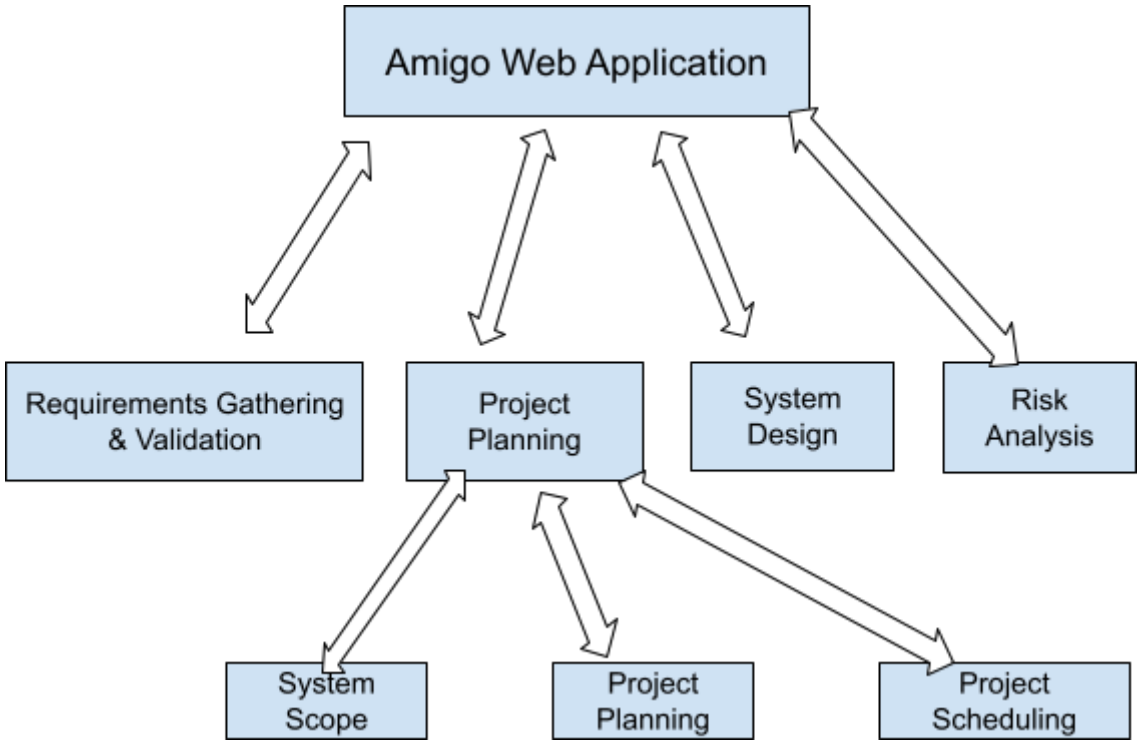

Fig 3.8: Module Analysis Breakdown

3.4 MODULE TRAINING

For our model, we use machine learning algorithms which give the prediction of the people suffering from stress, anxiety and depression. For this, we use the machine learning classification algorithms like Logistic Regression, Random Forest Tree, Support Vector Machine, Ada Boost, XGBoost, and Gaussian Naive Bayes in python programming language using Google Collab. We use the scikit-learn library to perform different algorithms accuracy, result, F1 score and confusion matrix.

```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn import tree
from sklearn import svm
from sklearn.naive_bayes import GaussianNB
from sklearn.ensemble import AdaBoostClassifier
from xgboost import XGBClassifier
```

Fig 3.9: Import Models

For the training and testing of our model, we have divided the transformed dataset in the ratio of 80:20. The shape of the training and the testing dataset is (31820, 21) and (7955, 221), respectively

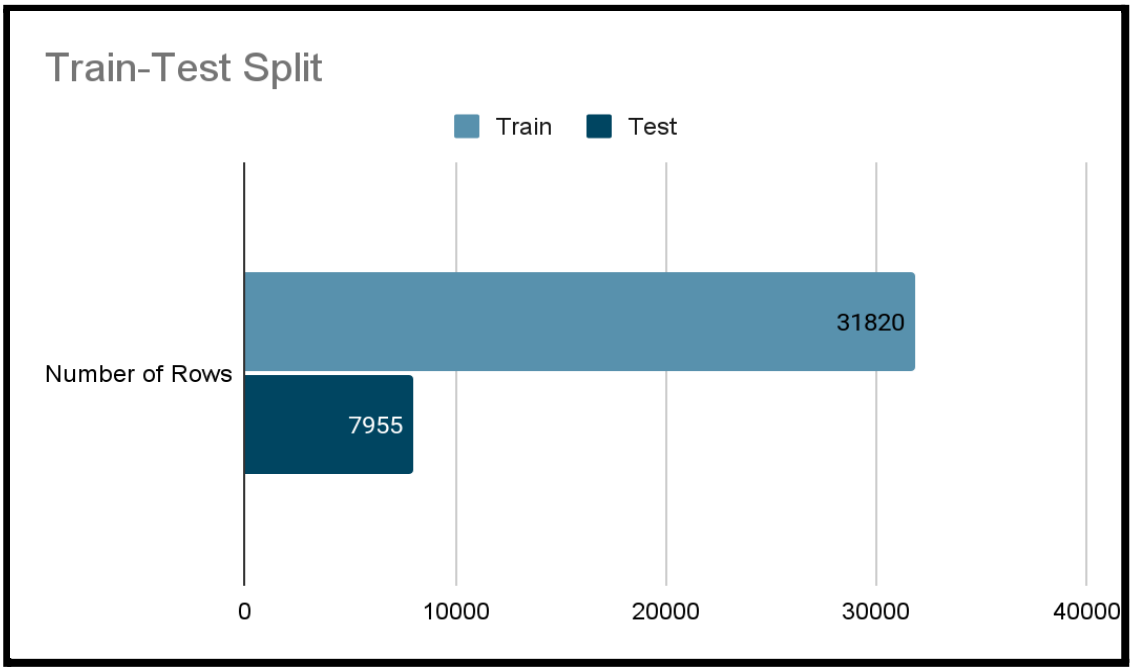

Fig 3.10: Train Test Split Distribution

3.4.1 Models Tested for Classification

3.4.1.1 Random Forest Tree:

It is a widespread machine learning algorithm and it also falls into the supervised learning technique. It is chosen for both Regression and Classification problems in machine learning. This algorithm is based on the notion of ensemble learning, which is a method of connecting numerous classifiers to crack a convoluted problem and enhance the rendition of the model.

```
[ ]  rf = RandomForestClassifier()
     rf.fit(x_train, y_train)
```

Fig 3.11: Random Forest Classifier

3.4.1.2 Logistic Regression:

It is the most commonly used machine learning algorithm for two classifications. Easy to implement and used as the baseline for any binary classification problem.

```
[ ]  lg = LogisticRegression()
     lg.fit(x_train, y_train)
```

Fig 3.12: Logistic Regression

3.4.1.3 Decision Tree:

It slips underneath the type ML (Supervised learning) technique, it can be used for both classification and Regression problems and it is mainly chosen for cracking classification problems.

```
[ ]  dcs = tree.DecisionTreeClassifier()
     dcs.fit(x_train,y_train)
```

Fig 3.13: Decision Tree Classifier

3.4.1.4 AdaBoost:

It is called Adaptive Boosting. It is a technique in ML seas as an Ensemble method. It used a decision tree with one level which means a decision tree with 1 split only.

```
abc = AdaBoostClassifier()
abc.fit(x_train,y_train)
```

Fig 3.14: Ada Boost Classifier

3.4.1.5 XGBoost:

It is an ensemble learning method. Periodically, it may not stand adequate to depend upon the outcomes of just one machine learning model. Ensemble learning proposes a systematic answer to merge the predictive ability of numerous learners. The consequent is a single model which presents the aggregated outcome from numerous models.

```
[ ] xg = XGBClassifier()
    xg.fit(x_train,y_train)
```

Fig 3.15: XG Boost Classifier

3.4.1.6 Support Vector Machine (SVM):

It is a supervised algorithm which is mostly used for the classification or Algorithm which is used to analyse regression and data for classification

```
svs = svm.SVC()
svs.fit(x_train,y_train)
```

Fig 3.16: Support Vector Machine

3.4.1.7 Gaussian Naive Bayes:

Gaussian Naive Bayes is a variant of Naive Bayes that follows Gaussian normal distribution and supports continuous data.

```
gb = GaussianNB()
gb.fit(x_train,y_train)
```

Fig 3.17: Gaussian Naive Bayes

3.5 WEBSITE REQUIREMENTS/DELIVERABLES:

3.5.1 Website Contents

1. Secure account creation with:
    i. Anonymous user accounts with anonymous names
    ii. User database for storing user information securely
2. Secure live chatroom with:
    i. Facility to create channels and invite users/friends
    ii. Send direct messages to other users
    iii. Ability to flag abusive or hate inducing messages
    iv. Facility of extensions like YouTube, emojis, and Giphy are provided
    v. Facilitate human moderation via giving the ability to ban disruptive users
3. Access to AI-powered questionnaire to track:
    i. Depression, Anxiety or Stress Score
    ii. Various Machine Learning models were analysed and compared and the highest performing model was used for classification

iii. The psychometric test DASS-42, Ten Item Personality Inventory and demographic questions are involved in the questionnaire to produce results [14, 15].

4. AI-powered moderation to track abusive and hate-inducing messages in group chat.

3.5.2 Website Features:

The web application has the following features:

3.5.2.1 Group Chat Feature:

1. For building the Amigo chat we have used the Stream React SDK.
2. The Stream React SDK has a React chat message component library, that makes it seamless to build a chat application and supports extensions from Giphy, YouTube video embeddings etc.
3. We can access the chat database in the Stream admin dashboard.
4. The dashboard is also helpful to create and delete channels, add therapist accounts and delete user accounts.
5. The stream also provides a mechanism to report a message that a user may find offensive, hurtful or irrelevant.
6. The messages will then be highlighted in the reported message section of the Stream dashboard.
7. The admin of the Amigo chat can then review the message and upon his judgement, he can temporarily or permanently ban or block the user.
8. If a therapist is offline when a user sends him a text, then an SMS will be sent to the mobile number of the therapist with the help of Twilio integration in the chat application.

3.5.2.2 DASS Quiz Feature:

1. The DASS quiz is made in React with Node backend.
2. The user has to start the quiz and answer questions related to his daily-life and lifestyle habits.
3. At the end of the quiz, the website shows the score and the likelihood of them having stress, depression or anxiety.
4. The user can save an image of the result and share it on his social media.

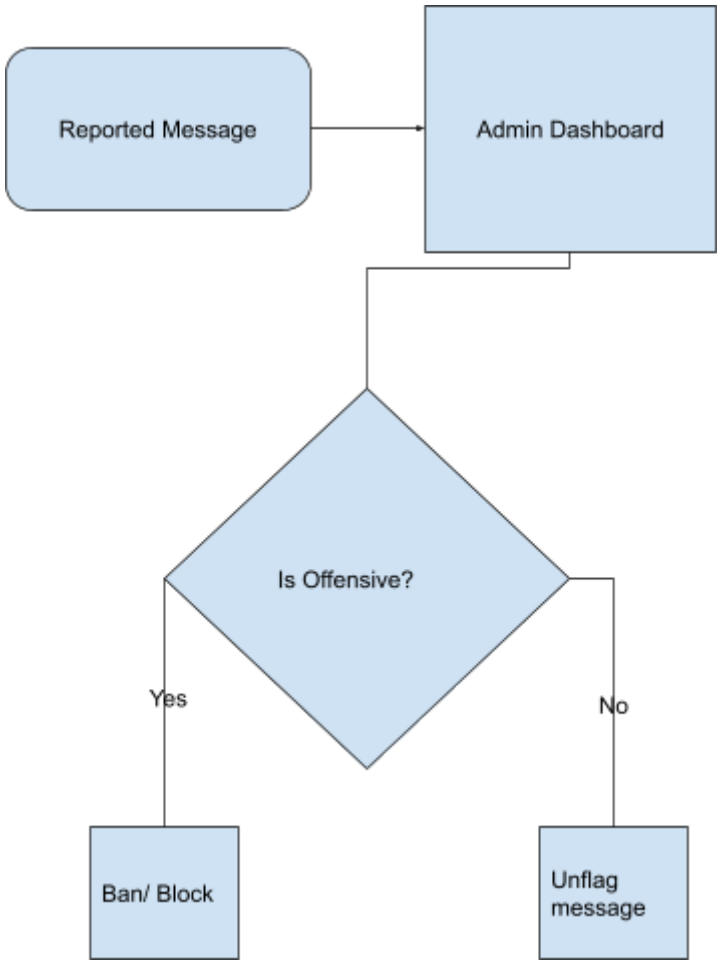

Fig 3.18: Message Report Workflow

CHAPTER 4

RESULT AND DISCUSSION

4.1 WEBSITE USER INTERFACE

The web application can be accessed through any browser on a desktop or mobile platform. The user needs to open the frontend client of the application hosted on Netlify at https://amigo-app.netlify.app/.

The following are the final User Interfaces of the various web pages of the website application.

4.1.1 Sign-Up Page

The Sign-Up window is the home page of the AMIGO web application.

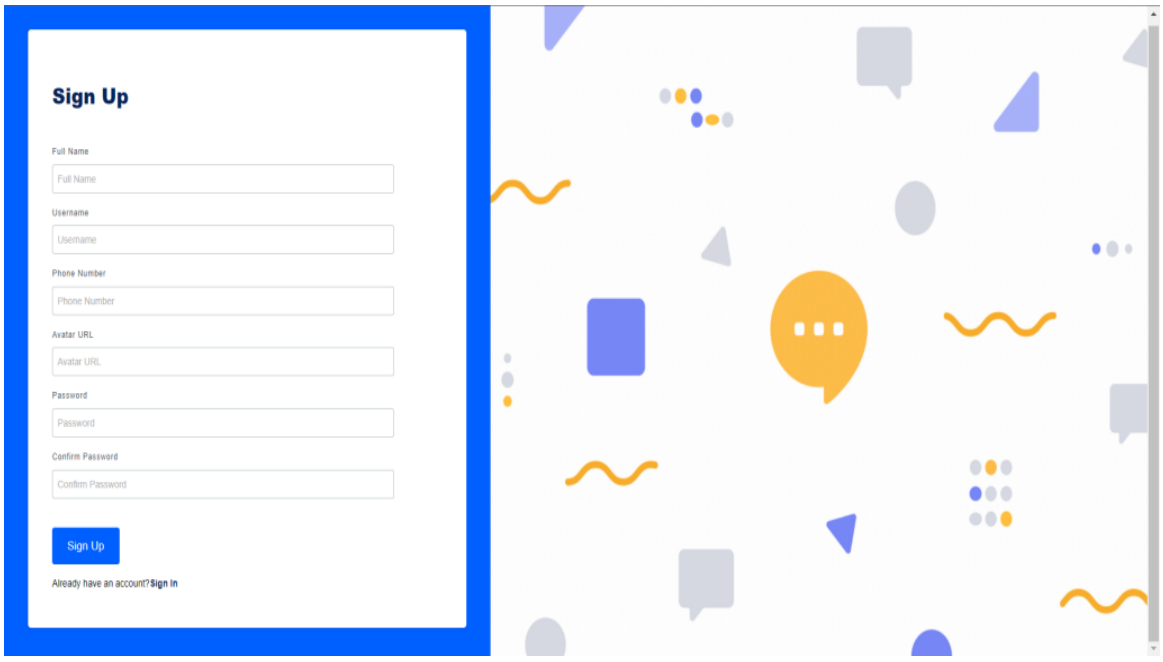

Fig 4.1: Sign-Up Page

4.1.2 Responsive Sign-Up Page

The front-end of the application is responsive, i.e., it can adapt to various screen sizes on different devices.

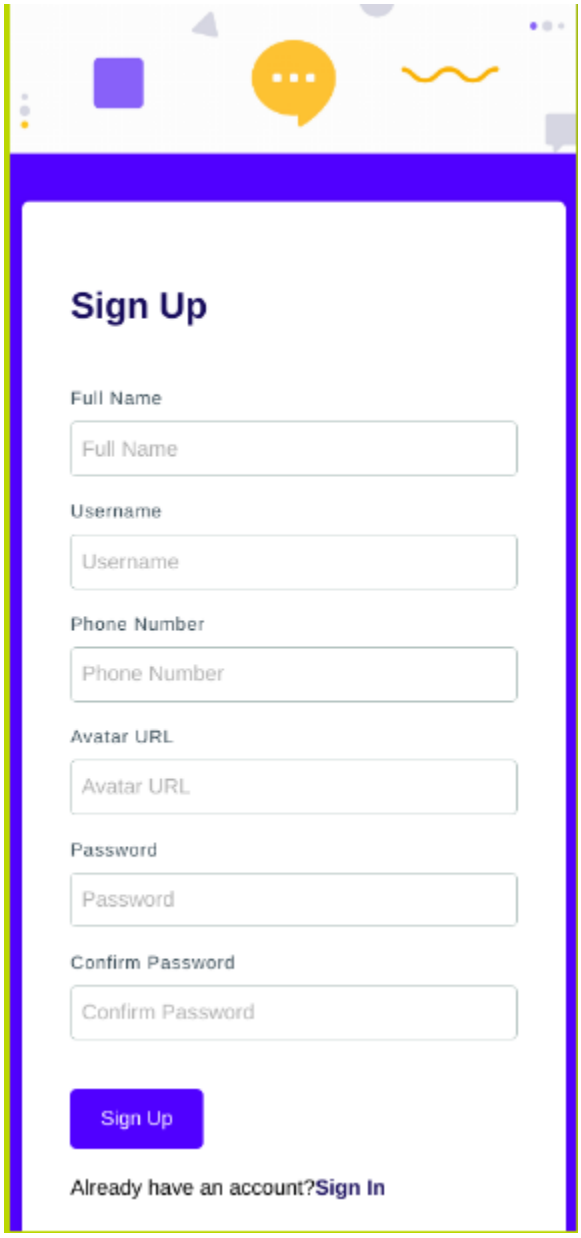

Fig 4.2: Responsive UI – Mobile Device Screen View

4.1.3 Sign-In Page

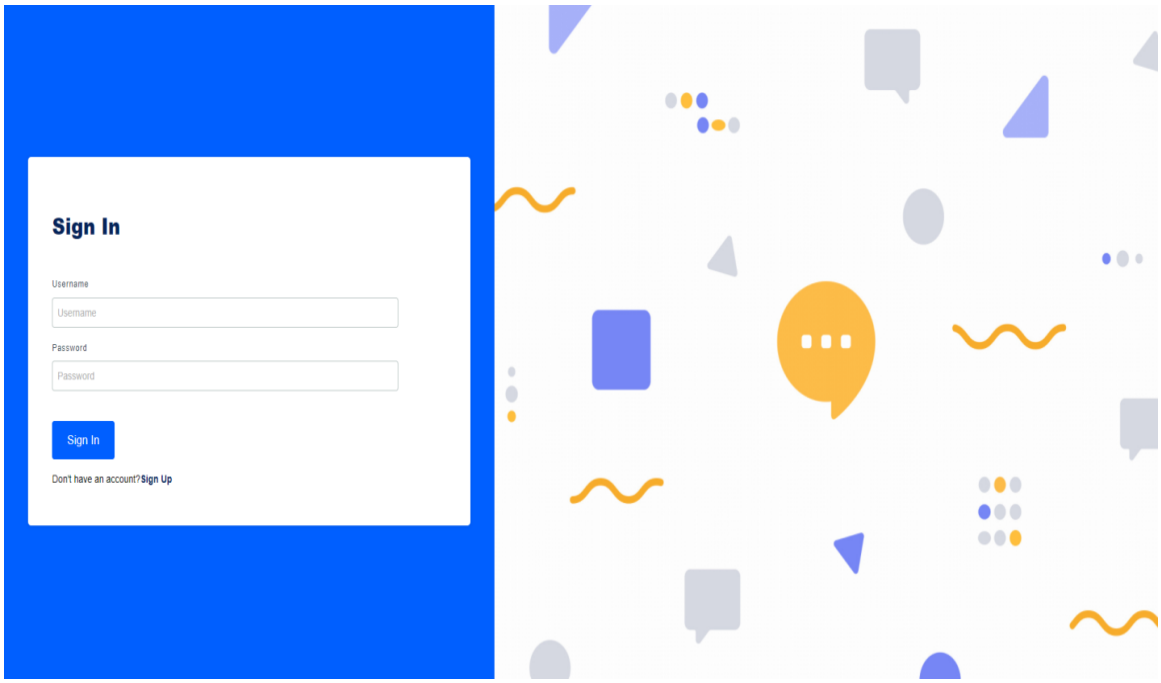

Fig 4.3: Sign-In Page

4.1.4 Chatroom Emotional Support Channel and Threads

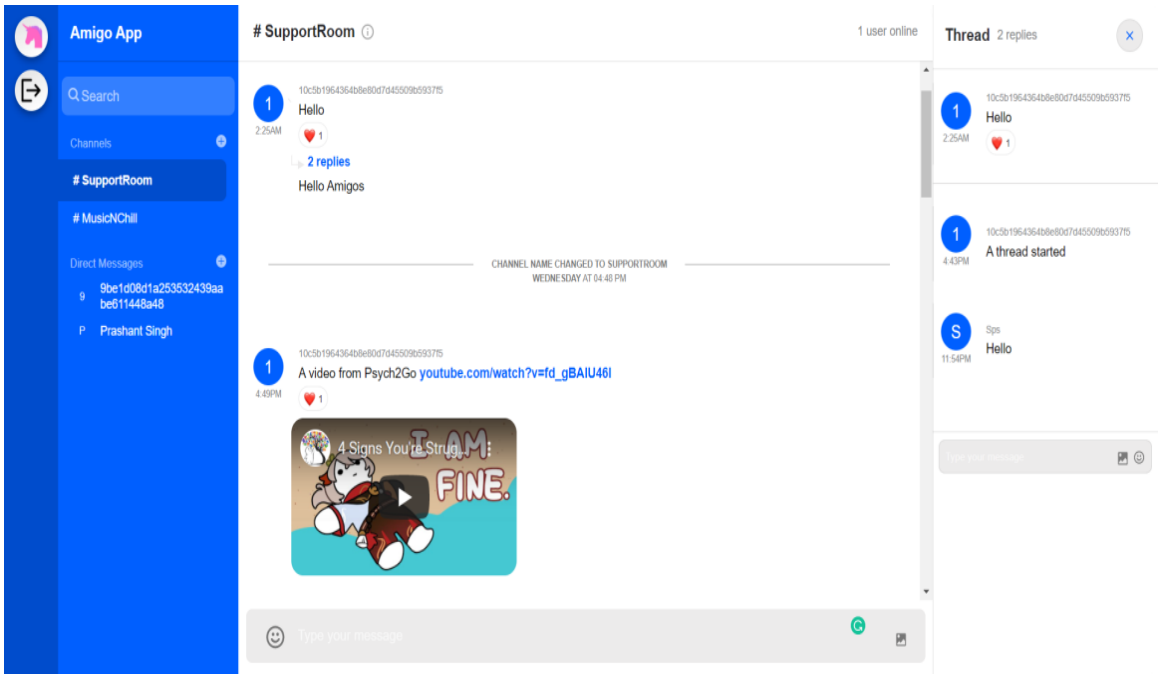

Fig 4.4: Chatroom Emotional Support Channel and Threads

4.1.5 YouTube Extension Support

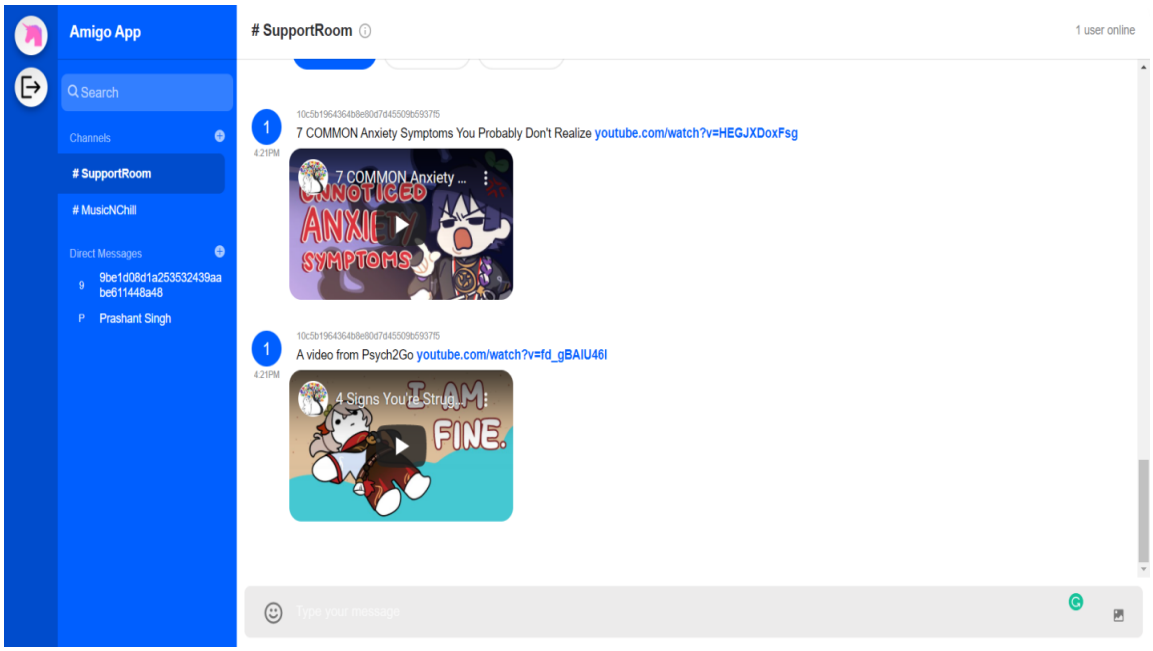

Fig 4.5: YouTube Extension Support

4.1.6 Emoji Support

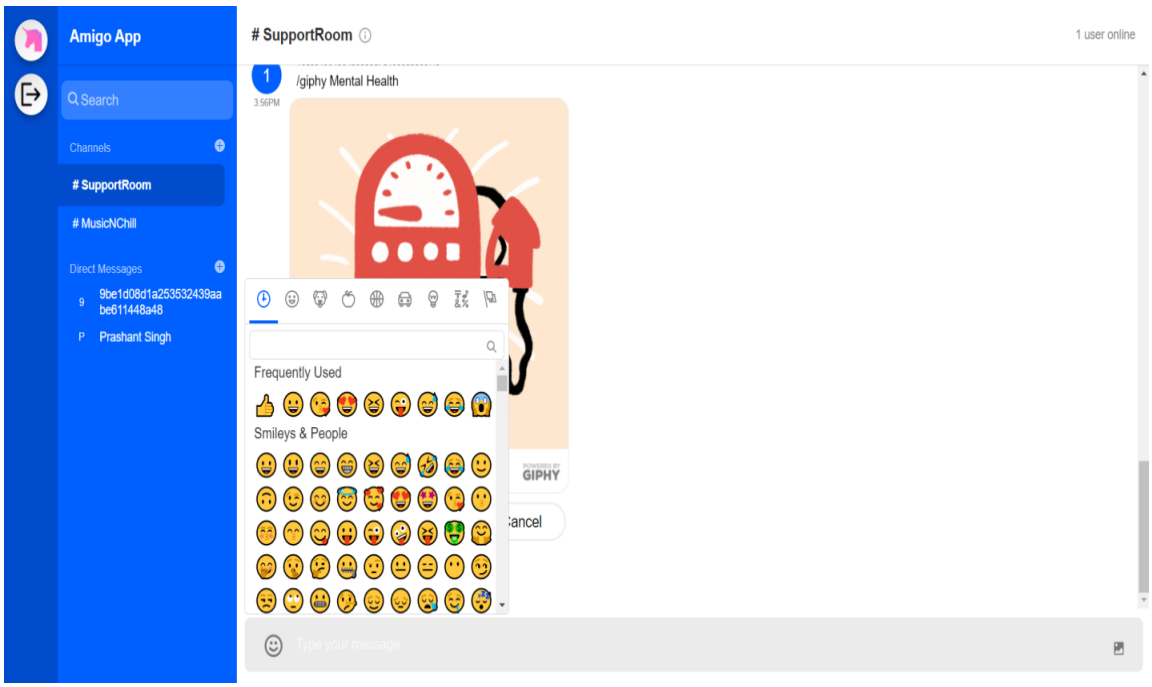

Fig 4.6: Emoji Support

4.1.7 Giphy Extension Support

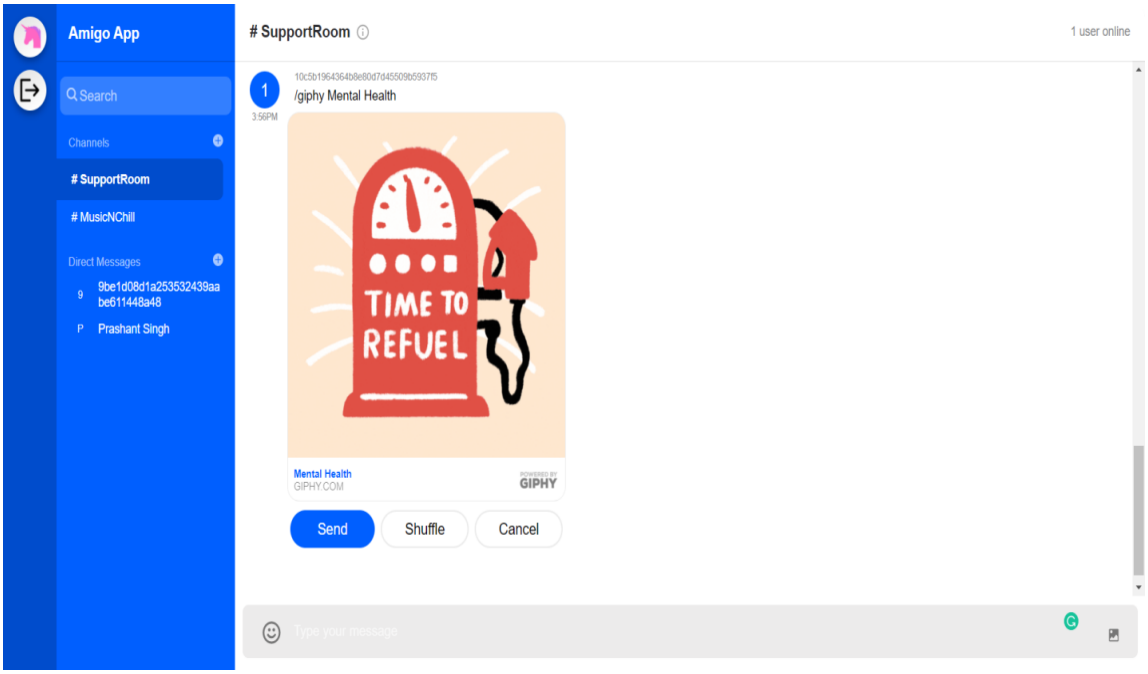

Fig. 4.7: Giphy Extension Support

4.2 MACHINE LEARNING CLASSIFIER

We analysed the performance of various machine learning algorithms in predicting the level of Depression, Stress or Anxiety in an individual. The dataset used for the research is made available by Kaggle Datasets Expert Yam Peleg which consists of data collected with an online version of the Depression Anxiety Stress Scales (DASS) [13].

The survey consisted of the Depression, Anxiety and Stress Scale (DASS-42) questionnaire. along with the time elapsed for each question and the order of occurrence in the survey.

The Ten Item Personality Inventory (TIPI) questionnaire, 16 validity-check word definition questions (out of which 3 were imaginary words). A few more questions were asked related to the individual's education, race, religion, gender, sexual orientation, age, native English speakers, marital status, family size, dominant hand, neighbourhood, major in university and voting status. Along with this on the server-side, the country ISO code, network location (for checking duplicity) and source of redirection were also recorded.

In the machine learning part of our project, we get an accuracy of 53% from Random forest Classifier, 46% from Logistic regression, 53% from Decision Tree, 53% from

Gaussian Naive Bayes, 53% from Support Vector Machine, 53% from AdaBoost, and 53% from XGBoost.

A more detailed overview of the various evaluation metrics is shown in table 4.1

Table 4.1: Evaluation metrics for different models implemented

Evaluation Matrix				
	Accuracy	**Precision**	**Recall**	**F1-score**
.RFC	53%	0.70	0.54	0.39
.LR	46%	0.29	0.47	0.36
DT	53%	0.70	0.54	0.39
GNB	53%	0.70	0.54	0.39
SVM	53%	0.70	0.54	0.39
ADA BOOST	53%	0.70	0.54	0.39
.XG BOOST	53%	0.70	0.54	0.39

The confusion matrices for the different classification models are as below. A Confusion matrix is an N x N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model. This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making [16]. Each row in a confusion matrix represents an actual class, while each column represents a predicted class.

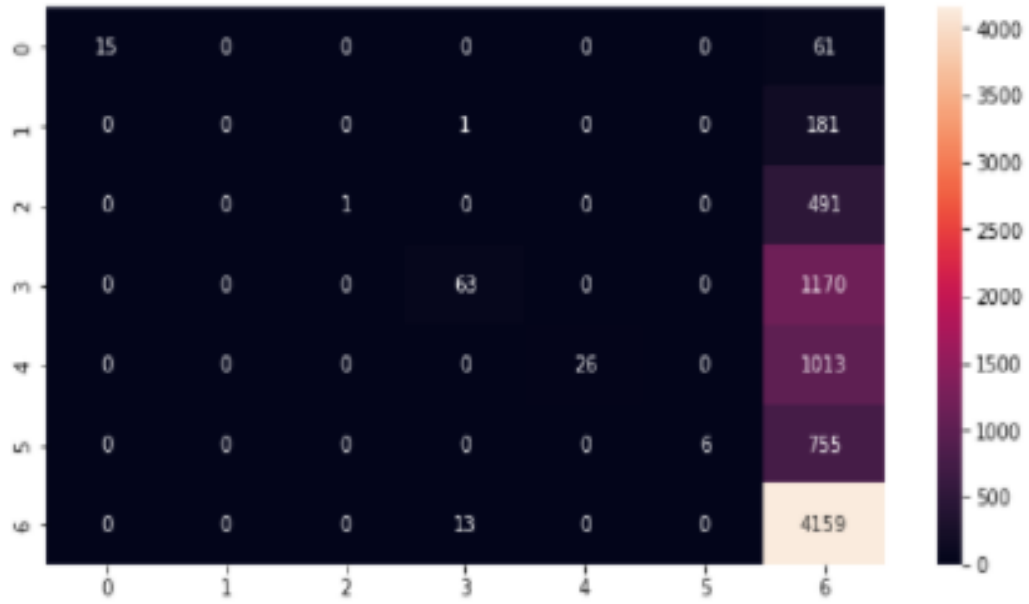

Fig 4.8: Confusion Matrix of Random Forest Classifier

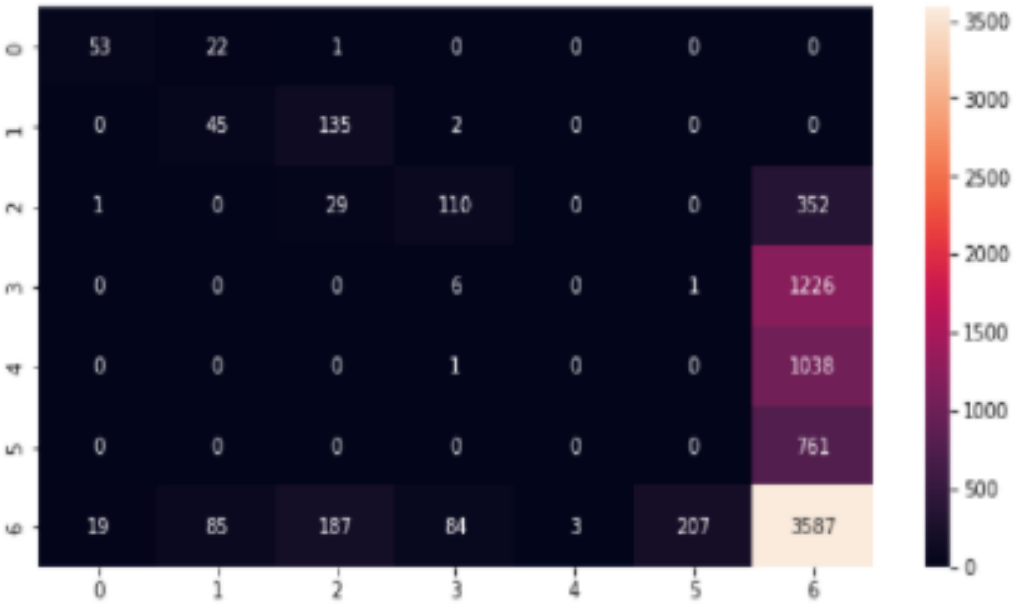

Fig 4.9: Confusion Matrix of Logistic Regression model

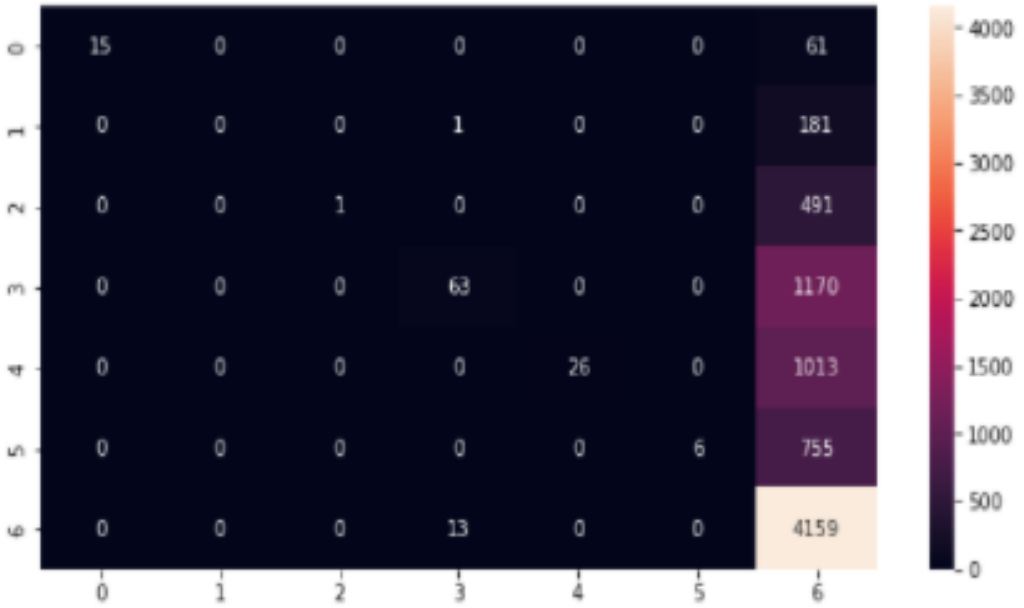

Fig 4.10: Confusion Matrix of Decision Tree Classifier

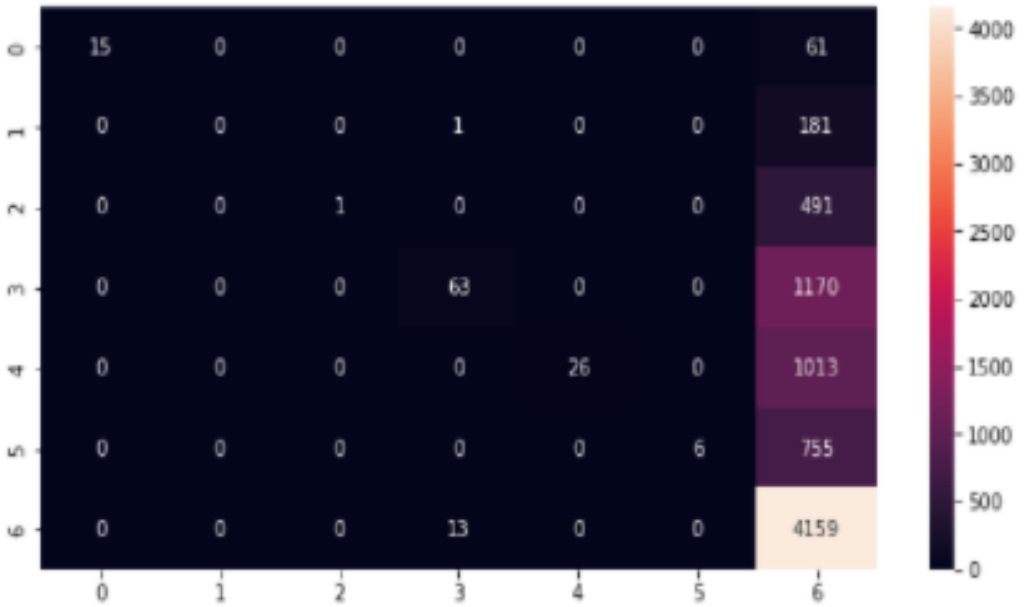

Fig 4.11: Confusion Matrix of Gaussian Naïve Bayes

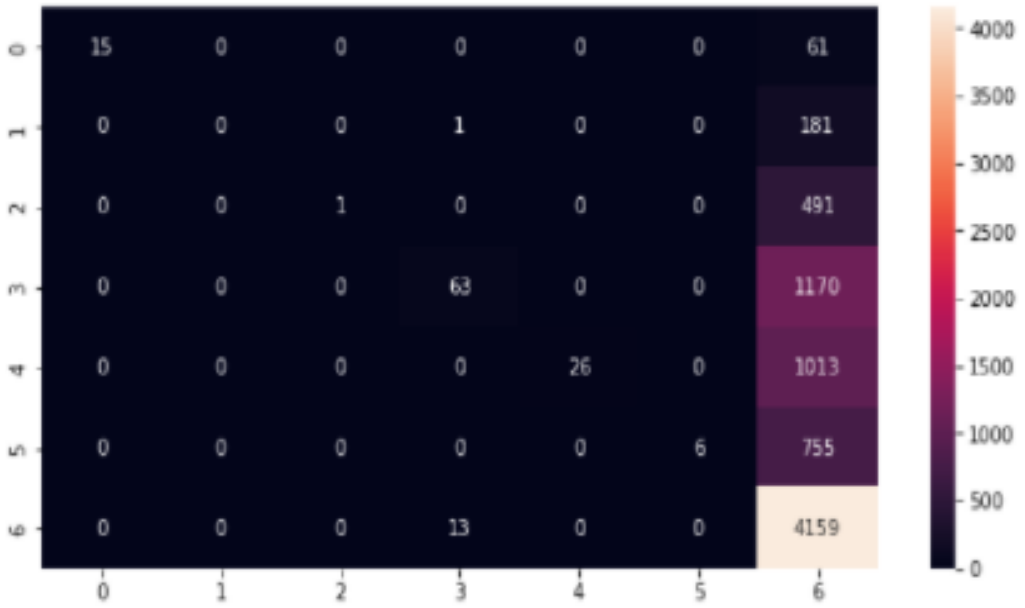

Fig 4.12: Confusion Matrix of Support Vector Machine

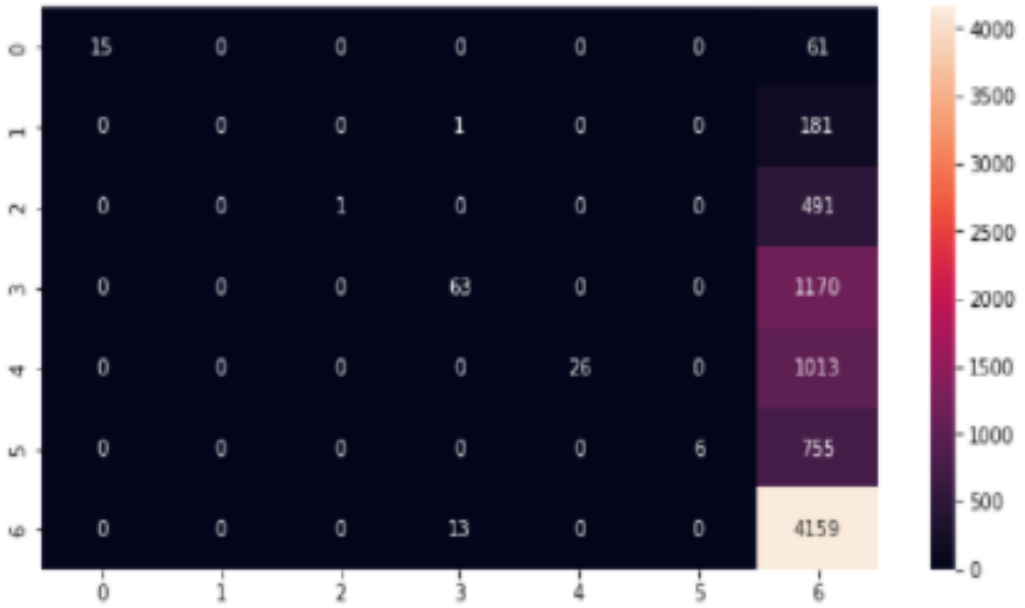

Fig 4.13: Confusion Matrix of Ada Boost

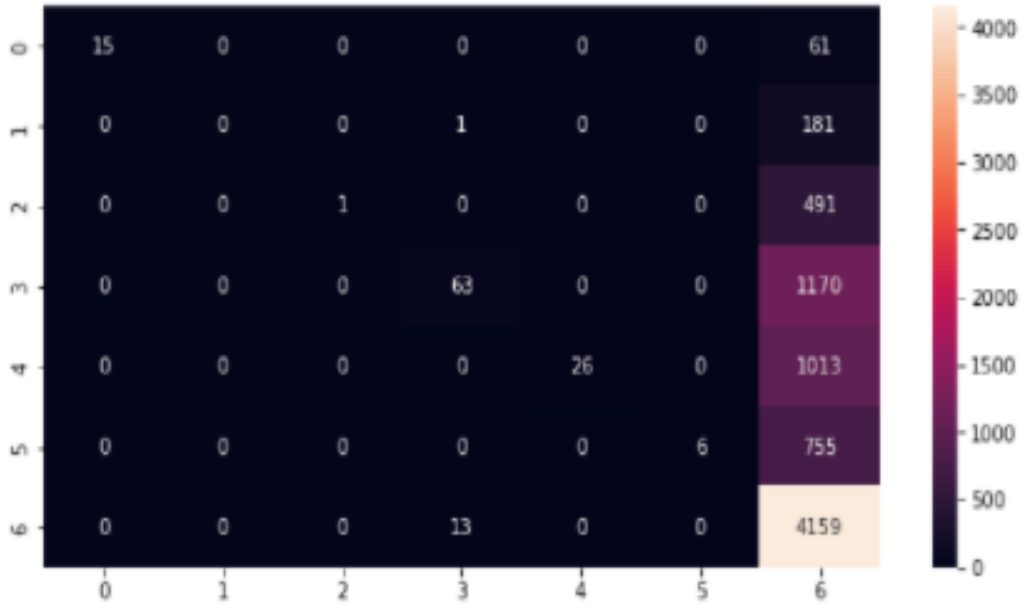

Fig 4.14: Confusion Matrix of XG boost

REFERENCES

[1] Fight depression, from https://fightdepression.quora.com

[2] Everything You Want to Know About Depression And How to ..., from https://mumuhealth.com/everything-you-want-to-know-about-depression/

[3] Social factors affecting old age -, from https://www.slideshare.net/Drraveesoni/social-factors-affecting-old-age

[4] Loneliness is a disease, from https://pjn.sbvjournals.com/doi/PJN/pdf/10.5005/pjn-7-1-18.

[5] David Bakker, Nikolaos Kazantzis, Debra Rickwood, Nikki Rickard, A randomized controlled trial of three smartphone apps for enhancing public mental health, 2018.

[6] David Bakker, Nikolaos Kazantzis, Debra Rickwood, Nikki Rickard, Development and Pilot Evaluation of Smartphone-Delivered Cognitive Behavior Therapy Strategies for Mood- and Anxiety-Related Problems: MoodMission, 2018.

[7] Aizenstros, A., Bakker, D., Hofmann, S., Curtiss, J., & Kazantzis, N., Engagement with smartphone-delivered behavioral activation interventions: A study of the MoodMission smartphone application. 2021.

[8] Tyshchenko, Y., "Depression and anxiety detection from blog posts data, 2018.

[9] Sau, A., Bhakta, I., Screening of anxiety and depression among the seafarers using machine learning technology, 2018.

[10] Reece, A. G., Reagan, A. J., Lix, K. L. M., Dodds, P. S., Danforth, C. M., Langer, E. J., Forecasting the Onset and Course of Mental Illness with Twitter Data, 2016.

[11] Braithwaite, S. R., Giraud-Carrier, C., West, J., Barnes, M. D., Hanson, C.L., Validating machine learning algorithms for Twitter data against established measures of suicidality, 2016.

[12] Du, J., Zhang, Y., Luo, J., Jia, Y., Wei, Q., Tao, C., Xu, H., Extracting psychiatric stressors for suicide from social media using deep learning, 2018.

[13] http://www2.psy.unsw.edu.au/dass/

[14] Lovibond, S.H., Lovibond, P.F., Manual for the Depression Anxiety Stress Scales (2nd ed.). Sydney: Psychology Foundation, 1995

[15] Crawford, J. R., & Henry, J. D., The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample, 2003.

[16] Everything you Should Know about Confusion Matrix for Machine Learning, from https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/