AUTOMATED DEPRESSION PREDICTION USING NLP

A PROJECT REPORT

Submitted by

ARCHANA K	**211419104016**
SWETHA R	**211419104285**
THARINI ABHINAYA S R	**211419104289**

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

APRIL 2023

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report **"AUTOMATED DEPRESSION PREDICTION USING NLP"** is the bonafide work of "**ARCHANA K (211419104016), SWETHA R (211419104285), THARINI ABHINAYA S R (211419104285)**" who carried out the project work under **Dr. T. TAMILVIZHI** supervision.

SIGNATURE

Dr. JABASHEELA,M.E., Ph.D.,
HEAD OF THE DEPARTMENT

DEPARTMENT OF CSE,
PANIMALAR ENGINEERING COLLEGE,
NASARATHPETTAI,
POONAMALLEE,
CHENNAI-600 123.

SIGNATURE

Dr. T. TAMILVIZHI,M.Tech., Ph.D.,
SUPERVISOR
ASSOCIATE PROFESSOR

DEPARTMENT OF CSE,
PANIMALAR ENGINEERING COLLEGE,
NASARATHPETTAI,
POONAMALLEE,
CHENNAI-600 123.

Certified that the above candidate(s) were examined in the Mini Project Viva-Voce Examination held on.....

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION BY THE STUDENT

We, **ARCHANA K (211419104016), SWETHA R (211419104285), THARINI ABHINAYA S R (211419104289)** hereby declare that this project report titled **"AUTOMATED DEPRESSION PREDICTION USING NLP"**, under the guidance of **Dr. T. TAMILVIZHI., M.Tech., Ph.D.,** is the original work done by us and we have not plagiarized or submitted to any other degree in any university by us.

1. **ARCHANA K**

2. **SWETHA R**

3. **THARINI ABHINAYA S R**

ACKNOWLEDGEMENT

We express our deep gratitude to our respected Secretary and Correspondent **Dr.P.CHINNADURAI, M.A, Ph.D..,** for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We would like to extend our heartfelt and sincere thanks to our Directors **Tmt. C.VIJAYARAJESWARI**, **Dr. C.SAKTHIKUMAR, M.E., Ph.D.** and **Tmt. SARANYASREE SAKTHIKUMAR, B.E., M.B.A.,** for providing us with the necessary facilities for completion of this project.

We also express our gratitude to our Principal **Dr. K.Mani, M.E., Ph.D..,** for his timely concern and encouragement provided to us throughout the course.

We thank the HOD of CSE Department, **Dr. JABASHEELA , M.E, Ph.D..,** for the support extended throughout the project.

We would like to thank my project guide **Dr. T. TAMILVIZHI, M.Tech., Ph.D.,** and all the faculty members of the Department of CSE for their advice and encouragement for the successful completion of the project.

ARCHANA K

SWETHA R

THARINI ABHINAYA S R

ABSTRACT

In today's fast-paced society, psychological health issues such as anxiety, depression, and stress have become prevalent among the general population. To address this problem, researchers have explored the use of machine learning algorithms to predict the likelihood of depression in individuals. As data sets related to depression become more abundant and machine learning technology advances, there is an opportunity to develop intelligent systems capable of identifying symptoms of depression in written material. By applying natural language processing and machine learning algorithms to analyze written text, such as social media posts, emails, and chat messages, researchers can potentially identify patterns and linguistic cues associated with depression. These patterns may include changes in word usage, tone, and sentiment. The dataset consists of text-based questions on this information channel. Machine learning Techniques are currently well suited for analysing the data and diagnosing the problem. The accuracy over the full attribute set and selected attribute set on various machine learning algorithms have been compared. In summary, while the potential for AI to aid in mental health diagnosis and treatment is exciting, it's important to proceed with care and consideration for the complexities of the field and the needs of patients.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
5.	**SYSTEM ARCHITECTURE**	
	5.1 Module Design Specification	33
	5.2 Algorithms	35
6.	**SYSTEM IMPLEMENTATION**	
	6.1 Client-side coding	41
	6.2 Server-side coding	45
7.	**SYSTEM TESTING**	
	7.1 Unit Testing	46
	7.2 Integration Testing	46
	7.3 Test Cases & Reports	47
8.	**CONCLUSION**	
	8.1 Results & Discussion	48
	8.2 Conclusion and Future Enhancements	48
	APPENDICES	
	A.1 Sample Screens	49
	REFERENCES	52

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

1. INTRODUCTION

1.1 OVERVIEW

In today's fast-paced society, psychological health issues such as anxiety, depression, and stress have become prevalent among the general population. To address this problem, researchers have explored the use of machine learning algorithms to predict the likelihood of depression in individuals. As data sets related to depression become more abundant and machine learning technology advances, there is an opportunity to develop intelligent systems capable of identifying symptoms of depression in written material. By applying natural language processing and machine learning algorithms to analyze written text, such as social media posts, emails, and chat messages, researchers can potentially identify patterns and linguistic cues associated with depression. These patterns may include changes in word usage, tone, and sentiment. The dataset consists of text-based questions on this information channel. Machine learning Techniques are currently well suited for analysing the data and diagnosing the problem. The accuracy over the full attribute set and selected attribute set on various machine learning algorithms have been compared. In summary, while the potential for AI to aid in mental health diagnosis and treatment is exciting, it's important to proceed with care and consideration for the complexities of the field and the needs of patients.

PROBLEM DEFINITION

The objective of this project is to predict depression in individuals using Natural Language Processing (NLP) techniques. Depression is a major mental health disorder that affects millions of people worldwide. NLP can be used to analyze various aspects of text, such as sentiment, tone, and word choice, to identify patterns that may indicate the presence of depression. The project aims to develop a model that can accurately predict depression in individuals using their written text, such as social media posts or chat messages. The model can be used to assist mental health professionals in identifying individuals who may be at risk of depression and provide them with appropriate care and support. The project will require a dataset of text samples from individuals with and without depression, as well as the development and testing of NLP algorithms to analyze the text.

CHAPTER 2

2. LITERATURE SURVEY

2.1 Title: [1]

Single classifier vs. ensemble machine learning approaches for mental health prediction.

Author name:

Jetli Chung and Jason Teo.

Year of publish:

2023

Description:

Early prediction of mental health issues among individuals is paramount for early diagnosis and treatment by mental health professionals. One of the promising approaches to achieving fully automated computer-based approaches for predicting mental health problems is via machine learning. As such, this study aims to empirically evaluate several popular machine learning algorithms in classifying and predicting mental health problems based on a given data set, both from a single classifier approach as well as an ensemble machine learning approach. The data set contains responses to a survey questionnaire that was conducted by Open Sourcing Mental Illness (OSMI). Machine learning algorithms investigated in this study include Logistic Regression, Gradient Boosting, Neural Networks, K-Nearest Neighbours, and Support Vector Machine, as well as an ensemble approach using these algorithms. Comparisons were also made against more recent machine learning approaches, namely Extreme Gradient Boosting and Deep Neural Networks. Overall, Gradient Boosting achieved the highest overall accuracy of 88.80% followed by Neural Networks with 88.00%. This was followed by Extreme Gradient Boosting and Deep Neural Networks at 87.20% and 86.40%, respectively. The ensemble classifier achieved 85.60% while the remaining classifiers achieved between 82.40 and 84.00%. The findings indicate that Gradient

Boosting provided the highest classification accuracy for this particular mental health bi-classification prediction task. In general, it was also demonstrated that the prediction results produced by all of the machine learning approaches studied here were able to achieve more than 80% accuracy, thereby indicating a highly promising approach for mental health professionals toward automated clinical diagnosis.

Methodology:

Predicting Mental Health: Machine learning algorithms include Logistic Regression, Gradient Boosting, Neural Networks, K-Nearest Neighbours, and Support Vector Machine, as well as an ensemble approach using these algorithms. Comparisons were also made against more recent machine learning approaches, namely Extreme Gradient Boosting and Deep Neural Networks.

2.2 Title: [2]

Designing a Tool to Address the Depression of Children During Online Education.

Author name:

Asma Alwadei and Reem Alnanih

Year of publish:

2022

Description:

Advances in communication and information technology have changed the way humans interact. During the COVID-19 pandemic, the technology for communication has caused depression and anxiety, including among children and teens. Depression among children and teens may go unrecognized and untreated, as parents and teachers may have difficulty recognizing the symptoms. COVID-19 has changed traditional learning methods, forcing children to stay home and connect through online education. Although some children may function reasonably well in less-structured environments, many children with significant depression suffer a noticeable change in social activities, loss of interest in an online school, poor online academic performance, or changes in appearance. Home quarantine has affected children's mental health, and it has become challenging for school counselors to predict depression in many children participating in online education. This study aims to design and develop a tool for predicting depression among children aged 7 to 9 years old by recording students' online classes and sending a note to the child's academic file. The idea of needing this tool arose as an output for applying the design thinking approach to the online education website during COVID-19. This inspired the authors to combine the lecture recordings and the prediction of depression into one tool. Image processing techniques are applied to generate the results predicted by the model on the collected videos. The overall accuracy for classifying depressed and not depressed videos is 89%.

Methodology:

Student's depression status: the student with no depression is labeled as 0, the student with depression is labeled as 1, and the empty videos are assigned with a value of -1. The videos are converted into frames and applied the concept of similarity measure points to standardize the number of frames for each video according to the minimum number of frames compared to all videos.

2.3 Title: [3]

Journaling Data for Daily PHQ-2 Depression Prediction and Forecasting.

Author name:

Alexander Kathan, Andreas Triantafyllopoulos, Xiangheng He, Manuel Milling and Tianhao Yan.

Year of publish:

2022

Description:

Digital health applications are becoming increasingly important for assessing and monitoring the wellbeing of people suffering from mental health conditions like depression. A common target of said applications is to predict the results of self-assessed Patient-Health-Questionnaires (PHQ), indicating current symptom severity of depressive individuals. Many of the currently available approaches to predict PHQ scores use passive data, e.g., from smartphones. However, there are several other scores and data besides PHQ, e.g., the Behavioral Activation for Depression Scale–Short Form (BADSSF), the Center for Epidemiologic Studies Depression Scale (CESD), or the Personality Dynamics Diary (PDD), all of which can be effortlessly collected on a daily. In this work, we explore the potential of using actively-collected data to predict and forecast daily PHQ-2 scores on a newly-collected longitudinal dataset. We obtain a best MAE of 1.417 for daily prediction of PHQ-2 scores, which specifically in the used dataset have a range of 0 to 12, using leave-one-subject-out cross-validation, as well as a best MAE of 1.914 for forecasting PHQ-2 scores using data from up to the last 7 days. This illustrates the additive value that can be obtained by incorporating actively-collected data in a depression monitoring application.

Methodology:

We experiment with four different regression models:

a) **XGBoost**

b) **Support Vector Machines**

c) **Random Forests**

d) **Multilayer Perceptrons (MLPs)**.

Each model comes with distinct associated hyper-parameters which is tuned using 3-fold nested-cross-validation on each train data set with non-disjoint subjects.

2.4 Title: **[4]**

Prediction of Public Mental Health by using Machine Learning.

Author name:

Ch.M.H. Saibaba, K V K Alekhya, K Yeshwanth and Praveen Tumuluru.

Year of publish:

2022

Description:

Interpretation public emotional wellness concerns utilizing data science and observing arrangements dependent on the discoveries from the data science tasks can be complicated and requires progressed methods, contrasted with regular data investigation projects. Have an extensive task the executive's interaction to guarantee that undertaking partners are capable and have sufficient information to carry out the data science procedure. Thus, this paper describes another structure that psychological wellness experts be able to use to address difficulties they realize utilizing data science. Albeit countless exploration papers have been distributed on open emotional well-being, few have tended to the utilization of data science in open psychological wellness. As of late, data science has altered the manner in which we oversee, investigate and influence data in medical care industry. High prevalence of mental health and the need for effective mental fitness care, blended with current advances in AI, has led to a growth in explorations of ways the sphere of system getting to know Machine Learning can assist inside the detection, prognosis and treatment of mental health issues.

Methodology:

Predicting with neural network for mental illness denoting all diagnosable mental disorders which are characterized by irregularities in reasoning, sentiments, or practices. Machine Learning can assist inside the detection, prognosis and treatment of mental health issues.

2.5 Title: **[5]**

The Impact of Mental Health due to Covid 19 – A Mental Health Detector Using Machine Learning.

Author name:

B Anishfathim, B Sreenithi, S Trisha, J Swathi and M Sindhu Priya

Year of publish:

2022

Description:

Stress has become a major part of human life during this Covid-19 pandemic due to the various internal and external expectations placed upon their shoulders. In this situation, Covid-19 is a very common and dangerous issue in the entire world, but as a result, people are going to notice the changes in their mental state such as depression, stress, and mood swings due to home quarantine. Thus, our paper will predict the stress levels using several algorithms of Machine Learning like K-Nearest Neighbour, Support Vector Machine, Naïve Bayes and Artificial Neural Networks. Data mining process is used to convert raw data into information for our purpose. Businesses can learn about their customers using software to look for the patterns for huge batches of data to develop effective strategies in marketing and cost reduction. Psychological health state plays a vital role in an individual's performance and characteristics. The expanding pace of life and the ways to focus is an important aspect of human life. Stress is often a sort of mental illness. There are possibilities of other problems such as heart attack, asthma, diabetes, etc. Consulting a psychiatrist is a traditional practice to overcome and manage any sort of mental disturbances. Since visiting a psychiatrist is very expensive, many people avoid the procedures. Our paper provides a technical approach with the help of a Machine Learning Model, which will determine the stress level of an individual, ultimately, eliminating the necessity of visiting psychiatrists. We will be analysing the strain at the beginning and finding out whether or not we can help them

in fighting their stress level. We have collected real-time data surveys from various regions based on questionnaires.

Methodology:

The following machine learning algorithms to predict the stress level during this quarantine.

Modules:

1. Acquisition of data.
2. Pre-processing the data.
3. Splitting the Training and Testing Data.
4. Classification of data.
5. Analysis on the performance of the model.

2.6 Title: [6]

Mental Health: Detection & Diagnosis.

Author name:

G. Parimala, R. Kayalvizhi and S. Nithiya

Year of publish:

2022

Description:

In this Work is built around Machine Learning Algorithms, Techniques, and Inferences. Machine learning is a data analysis methodology that allows for the automated creation of analytical models. It's a subset of AI that focuses on the idea that computers can learn from data, recognize patterns, and make decisions with little or no human intervention. We tested few ML algorithms depending upon our parameters that we would be further using to test our model. Thereby, the algorithm giving the best accuracy has been used. The front end hosted a questionnaire from where the data fetched is sent via API on our already trained model for the same. The results fetched from the model is reverted again by the API to show the above prediction. Besides this, we have created a 2-tier application because in case of a 1 tier application, in case of any small errors or changes too, the whole application must be changed. Mental Health is a very important part of us, controlling to a maj or extent our pattern to eat, sleep, reciprocate and express. Despite it being so important, it is neglected by most of us, primarily scared by acceptance and then the societal norms. According to a statistic released by the World Health Organization (WHO) in January 2020, approximately 264 million people of all age groups are suffering from depression which is a leading cause of disability around the world. Most people neglect thinking about their disorder which is caused by depression, and many a times land in severe conditions upon being noticed. As a result, technologies for early detection of mental illness and successful treatment are needed.

Methodology:

AI focuses on the idea that computers can learn from data, recognize patterns, and make decisions with little or no human intervention. Few ML algorithms are tested, depending upon our parameters that we would be further using to test our model. Thereby, the algorithm giving the best accuracy has been used.

2.7 Title: [7]

Predicting Mental Health Illness using Machine Learning Algorithm.

Author:

Konda Vaishnavi, Nikhitha Kamath, Ashwath Rao and N V Subba Reddy.

Year of Publish:

2021

Description:

Early detection of mental health issues allows specialists to treat them more effectively and it improves patient's quality of life. Mental health is about one's psychological, emotional, and social well-being. It affects the way how one thinks, feels, and acts. Mental health is very important at every stage of life, from childhood and adolescence through adulthood. This study identified five machine learning techniques and assessed their accuracy in identifying mental health issues using several accuracy criteria. The five machine learning techniques are Logistic Regression, K-NN Classifier, Decision Tree Classifier, Random Forest, and Stacking. We have compared these techniques and implemented them and also obtained the most accurate one in Stacking technique based with an accuracy of prediction 81.75%. Technological advancements such as smartphones, social media, neuroimaging, and wearables have enabled researchers of mental health and doctors to gather a tremendous amount of information at a rapid rate. Machine learning has developed as a reliable tool for analyzing these data. Machine Learning is the application of advanced probabilistic and statistical techniques to create computers that can learn from data on their own. This allows data patterns to be more easily and correctly discovered, as well as more accurate predictions from data sources. Natural language processing, speech recognition, computer vision, and Artificial intelligence have all benefited from machine learning, which allows developers and researchers to extract crucial data from datasets, deliver personalized experiences, and

develop intelligent systems. In domains like bioinformatics, ML has aided substantial progress by allowing for quick and scalable analysis of complicated data. Mental health data is also being investigated using similar analytic tools, with the potential to improve patient outcomes as well as improve understanding of psychological diseases and their management.

Methodology:

Five machine learning techniques and assessed their accuracy in identifying mental health issues using several accuracy criteria. The five machine learning techniques are Logistic Regression, K-NN Classifier, Decision Tree Classifier, Random Forest, and Stacking.

2.8 Title: [8]

Fuzzy Explainable Attention-based Deep Active Learning on Mental-Health Data.

Author:

Usman Ahmed, Jerry Chun-Wei Lin, and Gautam Srivastava.

Year of Publish:

2021

Description:

Abstract—In this paper, we propose a fuzzy classification deep attention-based model that expands emotional lexicons by using linguistic properties of actual patient authored texts. The active learning methods can expand the trained dataset and fuzzy rules over some time. As a result, the model itself can reduce its labeling efforts for mental health application. Thus, the designed model can solve issues related to vocabulary sizes per class, data sources, methods of creation, and create a baseline for human performance levels. This paper also gives fuzzy explainability by visualizing weighted words. Our proposed method uses a similarity-based method that includes a subset of unstructured data as the training set. Next, using an active learning mechanism cycle, our method updates the training model using new training points. This cycle is repeatedly performed until an optimal solution is reached. The designed model also converts all unlabeled texts into the training set. Our in-depth experimental results show that the emotion-based expansion enhances the testing accuracy and helps to build quality rules.

Methodology:

The standard PHQ-9 question-naire for authored text from patients. PHQ-9 is a standard procedure for measuring depression symptoms. The PHQ-9 method helps extract nine distinct behaviour types that incor-porate Diagnostic and Statistical Manual of Mental Disorders 5 (DSM-V)2.

2.9 Title: [9]

Predicting the Utilization of Mental Health Treatment with Various Machine Learning Algorithms.

Author:

Meera Sharma, Sonok Mahapat, Adeethyia Shankar.

Year of Publish:

2020

Description:

Mental health (the health of one's mind) is an integral component of the overall state of a person. The WHO formally defines an individual with a healthy mind as someone who "realizes [their] own abilities, can cope with the normal stresses of life, can work productively, and is able to make a contribution to [their] community". Someone with poor quality mental health is far more susceptible to mental disorders, either occasional or chronic, which negatively impact one's emotions, mood, and overall behavior. Diagnosis and treatment of people with mental health disorders remains challenged partly due to the public stigma attached to mental health. People with mental illness are challenged by stereotypes and prejudice that result from misconceptions about mental illness while struggling with the symptoms and disabilities resulting from the illness. In recent times, artificial intelligence and big data have begun increasingly used in healthcare, especially in mental health. Data can be sourced from many different sources like providers and non-providers. The use of big data is only expected to grow in the coming years.

Methodology:

Decision Tree model is an information-mapping method with a tree-like structure commonly used for a variety of purposes in machine learning. There are multiple types of decision trees, but the one we used on our research is a Classification and Regression

Tree, commonly abbreviated as CART, to conduct binary classification on our data

2.10 Title: [10]

Machine Learning Based Solutions for Real-Time Stress Monitoring.

Author:

Rajdeep Kumar Nath, Himanshu Thapliyal, Allison Caban-Holt and Saraju P. Mohanty.

Year of Publish:

2020

Description:

Stress is the reaction to adverse environmental situations that challenges the typical adaptive capability as perceived by an individual. Although positive stress (eustress) helps the individual to stay focused to deal with adversities, negative stress (distress) causes the activation of the HPA (hypothalamic-pituitary-adrenocortical) axis. Prolonged activation of the HPA axis may cause physiological and psychological disorders. Psychological stress is also found to affect physiological processes and has a negative effect on daily work performance and is thought to affect the national economy. Monitoring negative stress levels can provide useful information for identifying the stressors and provide an opportunity to adopt necessary precautions in preventing resulting disruption. Stress may be defined as the reaction of the body to regulate itself to changes within the environment through mental, physical, or emotional responses. Recurrent episodes of acute stress can disturb the physical and mental stability of a person. This subsequently can have a negative effect on work performance and in the long term can increase the risk of physiological disorders like hypertension and psychological illness such as anxiety disorder. Psychological stress is a growing concern for the worldwide population across all age groups. A reliable, cost-efficient, acute stress detection system could enable its users to better monitor and manage their stress to mitigate its long-term negative effects. In this article, we will review and discuss the literature that has used machine learning based approaches for stress detection. We will also review the existing solutions in the

literature that have leveraged the concept of edge computing in providing a potential solution in real-time monitoring of stress.

Methodology:

They have used edge computing framework for real-time stress monitoring. The idea of edge computing is to migrate some of the computing capacity close to the end-point where data is collected also known as the edge. When computations are done at the edge, the response time from the system reduces considerably as the data does not need to transmit through a very long distance.

2.11 Title: [11]

ILIOU Machine Learning Preprocessing Method for Depression Type Prediction.

Author:

Theodoros Iliou, Georgia Konstantopoulou, Mandani Ntekouli And Christina Lymperopoulou.

Year of Publish:

2019

Description:

Depression in its various forms (major depression, dysthymia, depressive phase of bipolar disorder) is for modern society a widespread phenomenon that causes great disability and incapacity. It causes great pain, but also has high economic cost to affected individuals and society, mainly due to large consequences but also because cured in all walks of life (National Depressive and Manic - Depressive Association Consensus Statement on the Under treatment of Depression - US National Committee for the Depressive Disorder and Manic - Depressive Disorder, commonly accepted statement (Consensus) for suboptimal treatment of Depression [1]. Depression describes both a transient mood state experienced by virtually all individuals at some time in their life as well as a clinical or biobehavioral syndrome, usually called Major Depressive Disorder (MDD). The main objective of this study was to find a data preprocessing method to boost the prediction performance of the machine learning algorithms in datasets of mental patients. Specifically, the machine learning methods must have almost excellent classification results in patients with depression, in order to achieve the sooner the possible the appropriate treatment. In this paper, we establish ILIOU data preprocessing method for Depression type detection. The performance of ILIOU data preprocessing method and Principal Component Analysis preprocessing method was evaluated using the 10-fold cross validation method assessing seven Machine Learning classification algorithms, Nearest-neighbour classifier (IB1), C4.5 algorithm implementation (J48), Random Forest, Multilayer Perceptron (MLP),

Support Vector Machine (SMO), JRIP and Fuzzy Logic (FURIA), respectively. The classification results are presented and compared analytically. The experimental results reveal that the transformed dataset with new features after ILIOU preprocessing method implementation to the original dataset achieved 100% classification - prediction performance of the classification algorithms. So ILIOU data preprocessing method can be used for significantly boost classification algorithms performance in similar datasets and can be used for Depression type prediction.

Methodology:

There are 3 data preprocessing methods: PCA method, ILIOU preprocessing method and EVOLUTIONARY SEARCH method. Tested the initial dataset and the preprocessed datasets using 7 classification algorithms: the Nearest-neighbour classifier, J48, Random Forest, Multilayer Perceptron, Support Vector Machine, JRIP and FURIA.

CHAPTER 3
3. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

Mental stress is an increasingly common psychological issue leading diseases such as depression, addiction, and heart attack. In this study, an early detection framework based on electroencephalogram (EEG) data is developed for reducing the risk of these diseases. In existing frameworks, signals are often segmented into smaller sections prior to being input to a deep neural network. However, this approach ignores the fundamental nature of EEG signals as a carrier of valuable information (e.g., the integrity of frequency and phase, and temporal fluctuations of EEG components). A multiclass classification framework termed MuLHiTA was developed for the early detection of mental stress. The model effectively captured hierarchical temporal dependencies in intraslice and interslice regions of EEG signals using two complementary branches exhibiting attention mechanisms and BLSTM networks.

CONS:

- In the existing system, brain signals are used to find whether the person is in depression or not but it is a complex process since it uses wave signals.
- It is a physical device with complex mechanism and so, machine learning is not implemented.
- It is difficult to use in Telemedicine.

3.2 PROPOSED SYSTEM

- The proposed model is to build a machine learning model that is capable of the classifying whether the person is in depression or not.
- The depression is considered as the major widespread problem in the society and controlling them is very difficult. So, there is a greater chance for the people to get disturbed mentally.
- The machine learning is generally built to tackle these types of complicated task. It takes more amount of time to analyze these types of data manually.
- The machine learning can be used to classify the depression or not by using the previous data and make them to understand the pattern and improve the accuracy of the model by adjusting parameters and use that model as the classification model.
- Different algorithms can be compared and the best model can be used for classification purpose.

PROS:

- In the proposed system, machine learning will be implemented for more accurate results.
- NLP based depression classification will be predicted.
- This system can be implemented in TELEMEDICINE.
- Social Media Platforms – helps in analysing people's mental health.
- Deployed in a web browser, so that it retains the privacy of the patients.
- Brain signals are not required in this proposed system.
- It is a time efficient process.

3.3 FEASIBILITY STUDY

TECHNICAL FEASIBILITY:

For this project, the various technical resources such as dataset and the correct IDE's, GPUs for training the model, etc are available and are made use of them efficiently. The various libraries, packages and APIs are open-source and are easy to use to manipulate data.

SOCIAL FEASIBILITY:

This project helps the marketing officials to analyze their customers and suggest accordingly and can minimize time and cost.It can support the AI chatbots and respond to the customers accordingly.

3.4 HARDWARE ENVIRONMENT:

Processor	: intel i3
Hard Disk	: 80GB
RAM	: 2GB

3.5 SOFTWARE ENVIRONMENT:

Operating System	: Windows 10 or later
Tool	: Anaconda with Jupyter Notebook

LIBRARIES

1. Pandas:

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. In this project, pandas is used for preprocessing - i.e., cleaning the csv files.

2. NumPy:

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

3. MatPlot:

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

4. SciKit:

Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support-vector machines, DecisionTreeClassifier, CatBoost, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

CHAPTER 4

4. SYSTEM DESIGN

4.1 ER DIAGRAM

Entity relationship diagrams provide a visual starting point for database design that can also be used to help determine information system requirements throughout an organization. The below fig 4.1.1 depicts the entity relationship diagram.

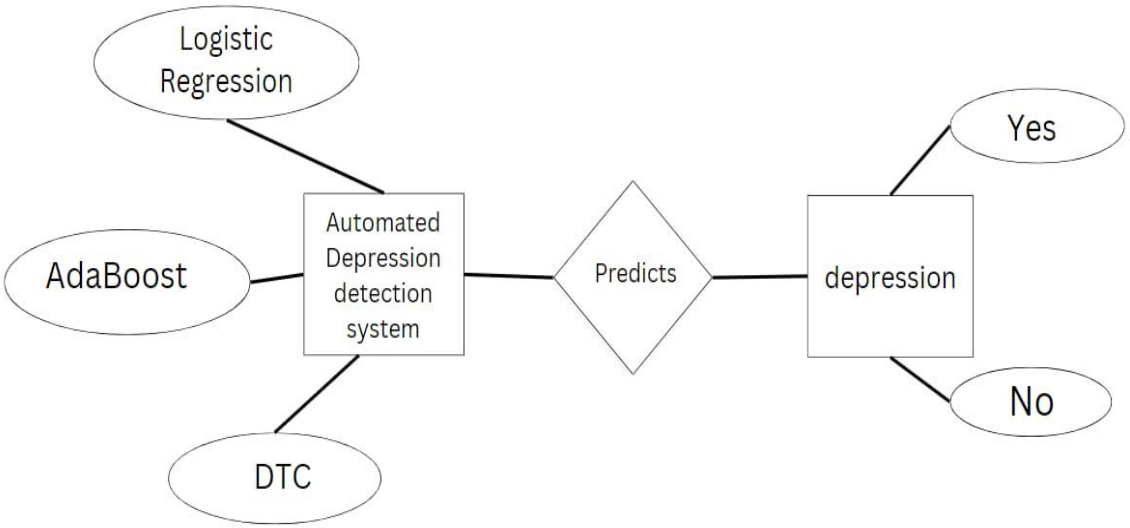

Fig.4.1.1 ER diagram

4.2 DATA DICTIONARY

This is normally represented as the data about the data stored in the database information system or as a part of a research project.

1. Words should be defined to understand what they need and not the variable need by which they may be described in the program.

2. Aliases are allowed when two or more entries show the same meaning.

3. A self-defining word should not be decomposed.

Data dictionary includes information such as the number of records in a file,the

frequency a process will run, security factors like pass word which the user must enter to get excess to the information.

4.2.1 INFORMATION TABLE

Attribute	Data type	Definition
Clean_text	varchar(50)	Textual data for training and testing
is_depression	int	In 0 and 1

Fig 4.2.1 Information table for the model

4.3 DATA FLOW DIAGRAM

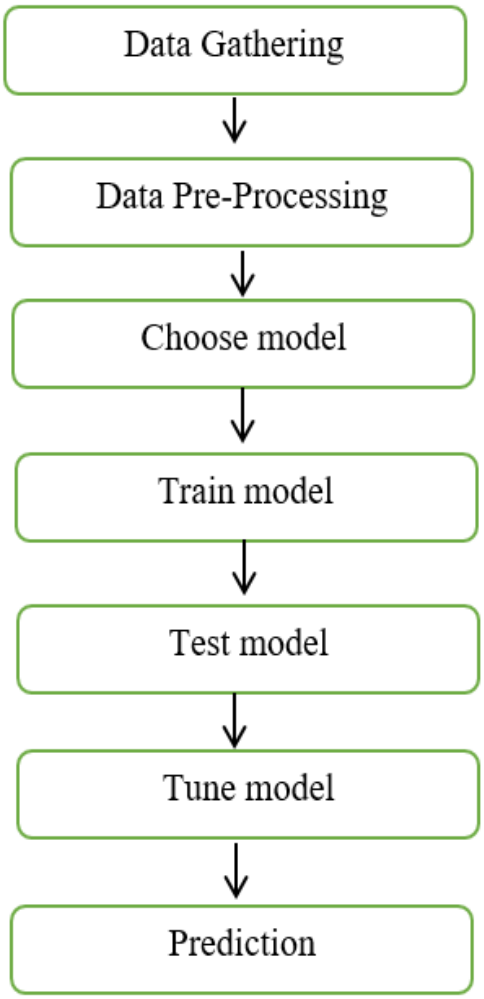

Fig.4.3.1 Data flow diagram

4.4 UML DIAGRAMS

USE CASE DIAGRAM

The below fig 4.5.1 depicts the operations performed by the system and the user. Use case

diagrams are considered for high level requirement analysis of a system. So when the requirements of a system are analyzed the functionalities are captured in use cases. So, it can say that uses cases are nothing but the system functionalities written in an organized manner.

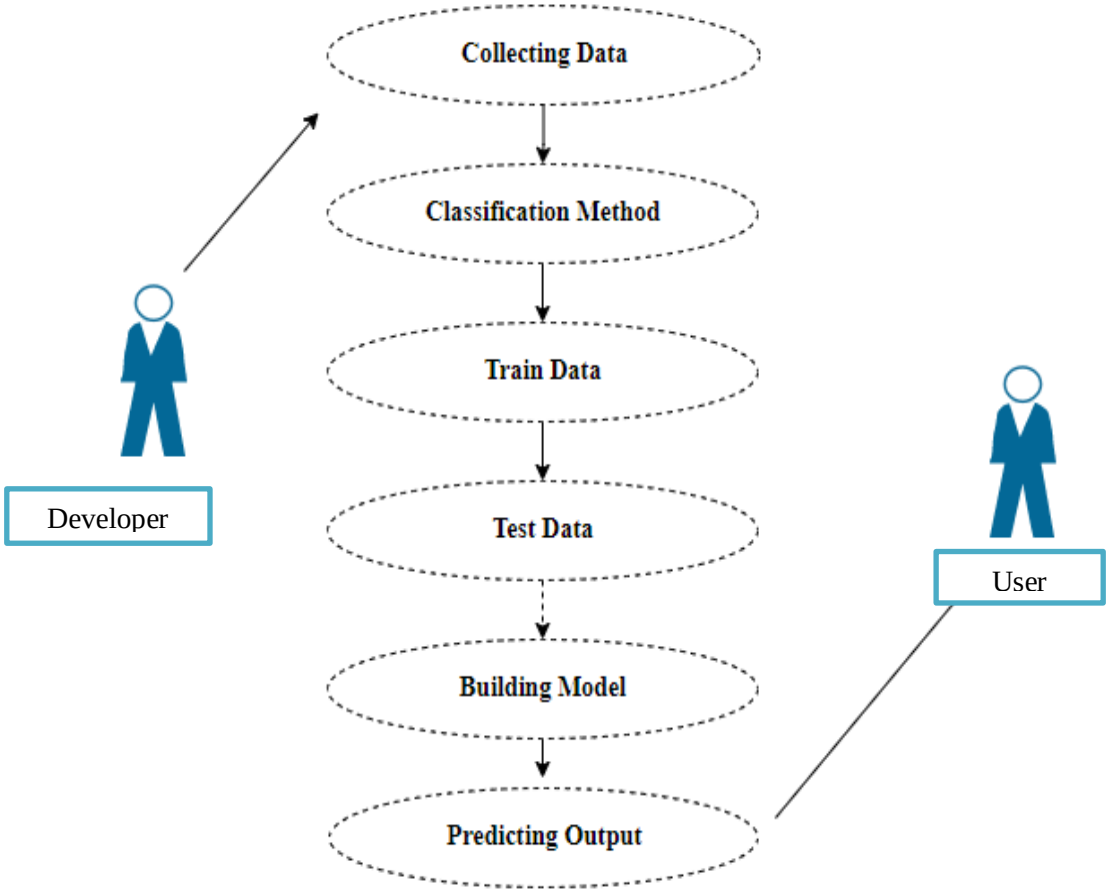

Fig 4.5.1 Use Case diagram

SEQUENCE DIAGRAM

Sequence diagrams model the flow of logic within your system in a visual manner, enabling you both to document and validate your logic, and are commonly used for both

analysis and design purposes. The below fig 4.5.4 depicts the sequence of activities happening during the prediction of depression when the user clicks the predict option in the website.

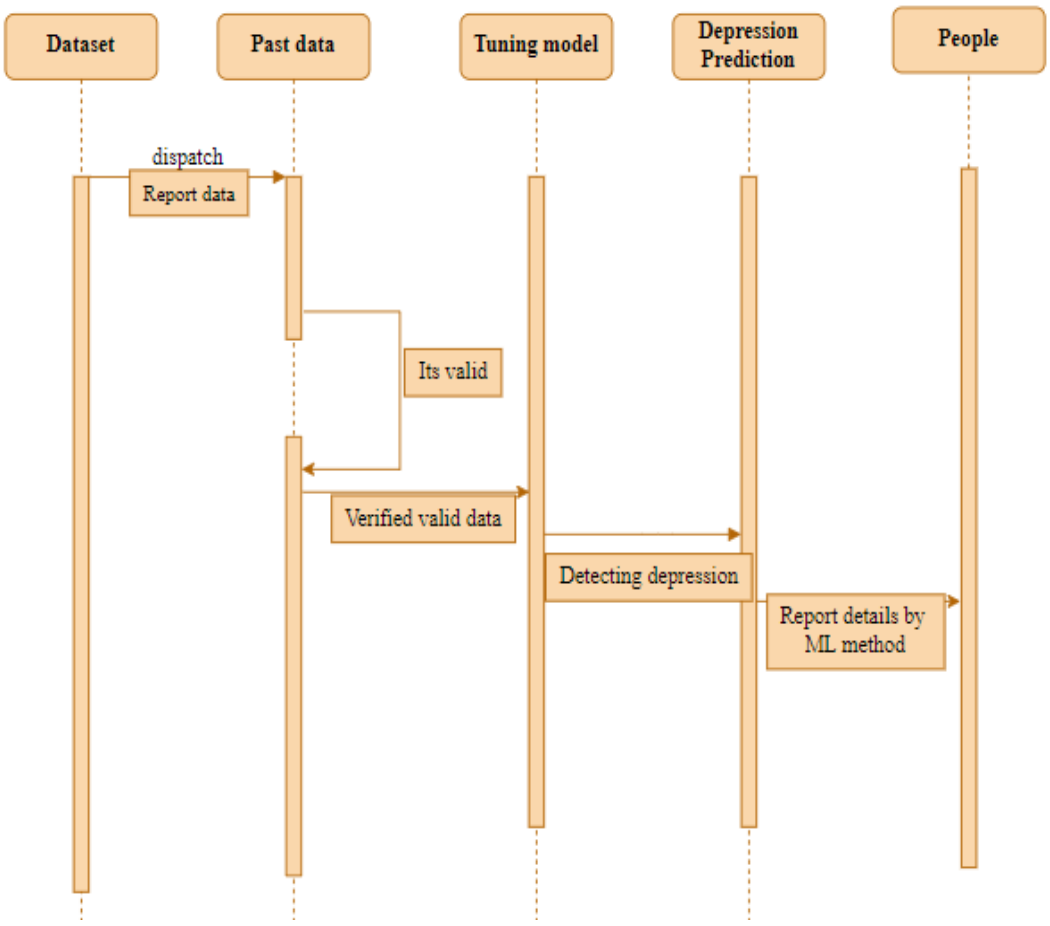

Fig 4.5.2 Sequence diagram

ACTIVITY DIAGRAM

Activity is a particular operation of the system. Activity diagrams are not only used for visualizing dynamic nature of a system but they are also used to construct the executable system by using forward and reverse engineering techniques. Activity diagram is some time considered as the flow chart. Although the diagrams looks like a flow chart but it is not. It shows different flow like parallel, branched, concurrent and single.

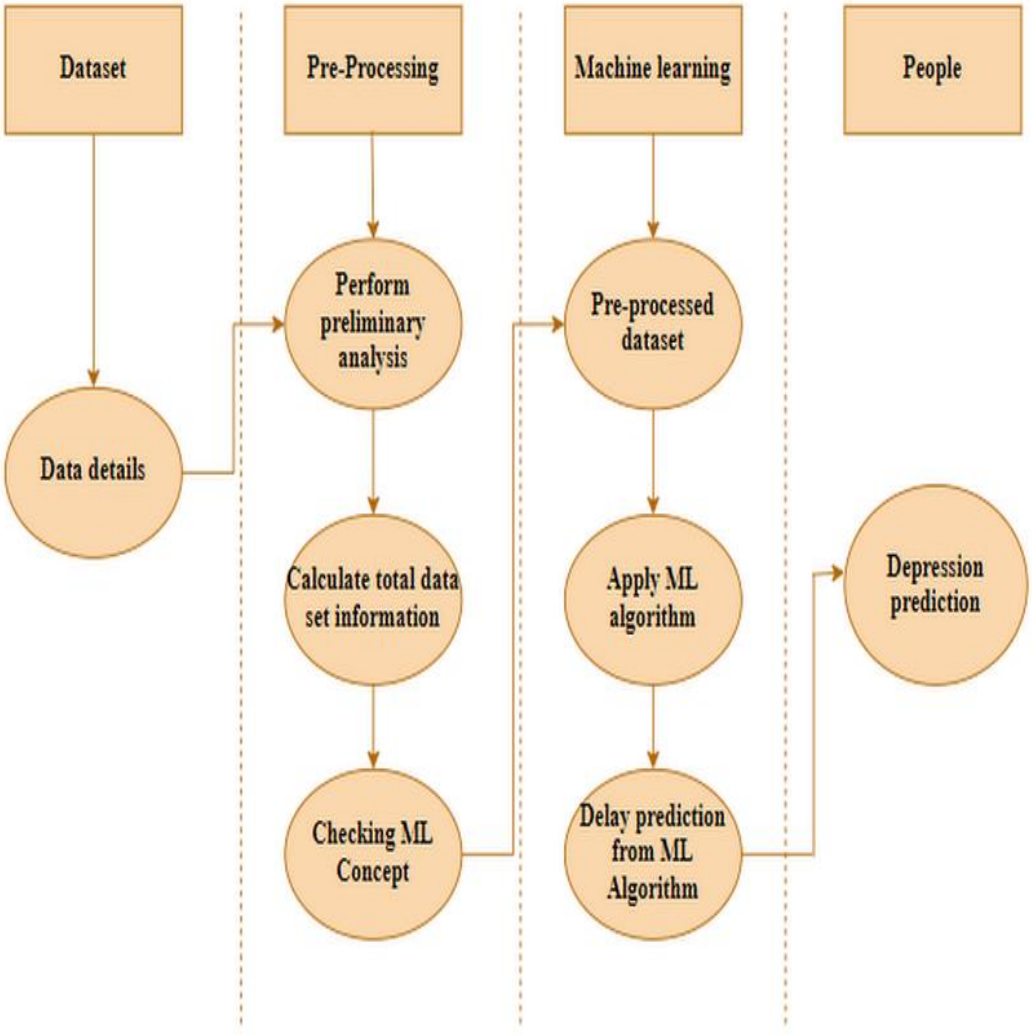

Fig 4.5.3 Activity diagram

CLASS DIAGRAM

The below fig 4.5.3 depicts the relationship between the user, interface and the model developed using the processes included as attributes.Both the training and testing data is used to develop the model.

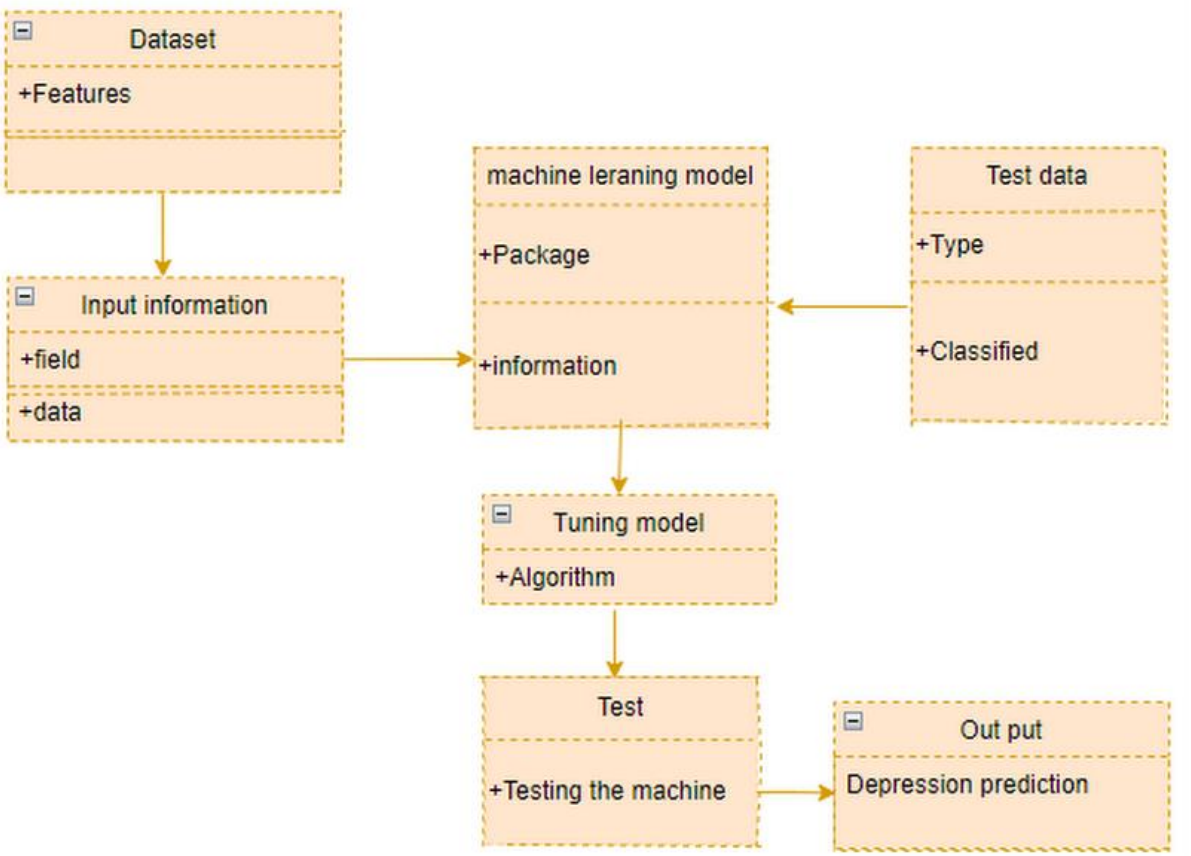

Fig 4.5.4 Class diagram

CHAPTER 5
SYSTEM ARCHITECTURE

5.1 MODULE DESIGN SPECIFICATION

- Data preprocessing
- Data Visualization
- Training and Evaluation
- Test set Prediction
- Deployment

DATA PREPROCESSING

Numpy and Pandas are commonly used Python libraries for data preparation and pre-processing. Numpy is used for mathematical operations and handling multi-dimensional arrays, while Pandas is used for data manipulation and analysis. Together, they provide a powerful toolkit for cleaning, transforming, and restructuring data before it is used for machine learning or other data analysis tasks. It involves transforming the data into a format that can be easily understood and processed by the algorithm. In Python, the following import statements are commonly used for pre-processing data:

import pandas as pd

import numpy as np

DATA VISUALIZATION

Data visualization is the representation of data in a graphical or pictorial format. It helps to identify patterns, trends, and relationships in the data. There are numerous libraries available in Python that enable the creation of visualizations, including but not limited to Matplotlib, Seaborn, Plotly, and Bokeh. These libraries provide a wide range of tools

and functions to create various types of plots, charts, and graphs including line plots, scatter plots, bar graphs, histograms, and heat maps. These visualizations can be customized with different colors, labels, titles, and annotations to make them more informative and appealing to the audience.

TRAINING AND EVALUATION

The developed model is then trained using the training dataset of RAVDESS so that the model is able to predict the emotion.

TESTING SET PREDICTION

It refers to the process of using a trained machine learning model to make predictions on the testing set. The goal is to see how well the model performs on data that it has not been trained on, which is a crucial step in assessing the model's overall accuracy and effectiveness. The model is tested using the testing dataset to analyze the efficiency and accuracy of the model.

DEPLOYMENT

In this module, the machine learning model that has been trained is typically saved in a pickle file format (.pkl file). This file can then be used for deployment in order to create a more user-friendly interface and provide accurate predictions for Depression Prediction.

5.2 ALGORITHMS

It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare multiple different machine learning algorithms in Python with scikit-learn. It can use this test harness as a template on your own machine learning problems and add more and different algorithms to compare. Each model will have different performance characteristics. Using resampling methods like cross validation, you can get an estimate for how accurate each model may be on unseen data. It needs to be able to use these estimates to choose one or two best models from the suite of models that you have created. When have a new dataset, it is a good idea to visualize the data using different techniques in order to look at the data from different perspectives. The same idea applies to model selection. You should use a number of different ways of looking at the estimated accuracy of your machine learning algorithms in order to choose the one or two to finalize. A way to do this is to use different visualization methods to show the average accuracy, variance and other properties of the distribution of model accuracies.

In the next section you will discover exactly how you can do that in Python with scikit-learn. The key to a fair comparison of machine learning algorithms is ensuring that each algorithm is evaluated in the same way on the same data and it can achieve this by forcing each algorithm to be evaluated on a consistent test harness.

Performance Metrics to calculate:

False Positives (FP): A person who will pay predicted as defaulter. When actual class is no and predicted class is yes. E.g. if actual class says this passenger did not survive but predicted class tells you that this passenger will survive.

False Negatives (FN): A person who default predicted as payer. When actual class is yes but predicted class in no. E.g. if actual class value indicates that this passenger survived and predicted class tells you that passenger will die.

True Positives (TP): A person who will not pay predicted as defaulter. These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes. E.g. if actual class value indicates that this passenger survived and predicted class tells you the same thing.

True Negatives (TN): A person who default predicted as payer. These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no. E.g. if actual class says this passenger did not survive and predicted class tells you the same thing.

True Positive Rate(TPR) = TP / (TP + FN)

False Positive rate(FPR) = FP / (FP + TN)

Accuracy: The Proportion of the total number of predictions that is correct otherwise overall how often the model predicts correctly defaulters and non-defaulters.

Accuracy calculation:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best.

Precision: The proportion of positive predictions that are actually correct.

Precision = TP / (TP + FP)

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labelled as survived, how many actually survived? High precision relates to the low false positive rate. We have got 0.788 precision which is pretty good.

Recall: The proportion of positive observed values correctly predicted. (The proportion of actual defaulters that the model will correctly predict)

Recall = TP / (TP + FN)

Recall(Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

General Formula:

F- Measure = 2TP / (2TP + FP + FN)

F1-Score Formula:

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

The below 3 different algorithms are compared:

- Decision Tree Classifier
- AdaBoost
- Logistic Regression

Decision Tree Classifier:

Introduction Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two

entities, namely decision nodes and leaves.

A tree has many analogies in real life, and turns out that it has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

```
from sklearn.tree import DecisionTreeClassifier

dt=DecisionTreeClassifier()
dt.fit(X_train,y_train)
```

ut[16]:
```
DecisionTreeClassifier()
```
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

n [17]:
```
# y_predict

y_predict=dt.predict(X_test)
```

Cat boost Classifier:

CatBoost is a machine learning algorithm that uses gradient boosting to achieve high accuracy in predictions for numerical, categorical, and textual data. It can be used for classification and regression tasks and is effective in handling noisy data. CatBoost natively handles categorical variables, offers faster training times than other boosting algorithms, and includes advanced techniques like ordered boosting and model introspection for better performance and interpretability.

```
In [14]: from sklearn.feature_extraction.text import CountVectorizer
         cv = CountVectorizer()
         cv_X = cv.fit_transform(X) # Fit the Data

In [15]: from sklearn.model_selection import train_test_split
         X_train, X_test, y_train, y_test = train_test_split(cv_X, y, test_size=0.2, random_state=42)

In [16]: from catboost import CatBoostClassifier
         cbc = CatBoostClassifier(verbose=0)
         cbc.fit(X_train,y_train)

Out[16]: <catboost.core.CatBoostClassifier at 0x2a96c518700>

In [17]: predict = cbc.predict(X_test)

In [18]: from sklearn.metrics import accuracy_score
         print('Accuracy of Catboost Classifier',accuracy_score(y_test,predict)*100)

         Accuracy of Catboost Classifier 95.15190691661279
```

Logistic Regression:

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes.

In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

Mathematically, a logistic regression model predicts P(Y=1) as a function of X. It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc.

Before diving into the implementation of logistic regression, we must be aware of the following assumptions about the same −

- In case of binary logistic regression, the target variables must be binary always and the desired outcome is represented by the factor level 1.
- There should not be any multi-collinearity in the model, which means the independent variables must be independent of each other.
- We must include meaningful variables in our model.
- We should choose a large sample size for logistic regression.

In [15]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(cv_X, y, test_size=0.2, random_state=42)
```

In [16]:
```
from sklearn.linear_model import LogisticRegression
clf = LogisticRegression()
clf.fit(X_train,y_train)
```

Out[16]:
```
LogisticRegression()
```
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [17]:
```
predict = clf.predict(X_test)
```

In [18]:
```
from sklearn.metrics import accuracy_score
print(accuracy_score(y_test,predict)*100)
```

CHAPTER 6

SYSTEM IMPLEMENTATION

6.1 CLIENT-SIDE CODING

Home.html

```
<!DOCTYPE html>
<html>
<head>
	<title>Home</title>
	<link rel="stylesheet" type="text/css" href="{{ url_for('static', filename='css/bootstrap.min.css') }}">
	<style>
		.back{
		  background-image: url("{{ url_for('static', filename='image/img2.jpg') }}");
		  background-repeat: no-repeat;
		  background-attachment: fixed;
		  background-size: 100% 100%;
		}
		.white{
		color:white;
		}
		.space{
		margin:10px 30px;
		padding:10px 10px;
		background: lightblue;
```

```
            width:500px
            }
            .gap{
            padding:10px 20px;
            }
            </style>
</head>
<body class="back">
        <header class="jumbotron" style="height:100px;">
            <div class="container">
            <center>
            <h2>AUTOMATED DEPRESSION PREDICTION USING NLP</h2>
            </center>
        </div>
        </header>
        <center>
        <div class="card ml-container" style="width:40%" >
            <form class="form-group" action="{{ url_for('predict')}}"
        method="POST">
            <label class="black"  for="">Enter the description here</label>
            <!-- <input type="text" name="comment"/> -->
            <textarea name="message" class="space form-control" rows="6"
cols="50"></textarea>
            <br/>
            <input type="submit" class="btn btn-success btn-block"
style="width:350px;padding:20px" value="Predict">
        </form>
```

```
        </div>
</center>
</body>
</html>
```

Result.html

```
<!DOCTYPE html>
<html>
<head>
        <title></title>
   <link rel="stylesheet" type="text/css" href="{{ url_for('static',
filename='css/bootstrap.min.css') }}">
        <style>
                .back{
                 background-image: url("{{ url_for('static', filename='image/img2.jpg')
}}");
                  background-repeat: no-repeat;
                  background-attachment: fixed;
                  background-size: 100% 100%;
                }
                center{
                padding-top:10%;
                }
                a{
                color:red;
                }
                </style>
</head>
```

```
<body class="back">
        <header class="jumbotron" style="height: 100px;">
                <div class="container">
                <h2 style="text-align:center">DEPRESSION PREDICTION USING
NLP</h2>
        </div>
        </header>
        <center>
        <div class="card" style="width:30%">
        <p style="color:red;font-size:20;text-align: center;"><b>Result for depression
prediction</b></p>
        <div class="results">
{% if prediction[0] == 'Depression' %}
        <h2 style="color:red;">PERSON IN DEPRESSION</h2>
        {% elif prediction[0] == 'Not in Depression' %}
        <h2 style="color:green;">PERSON NOT IN DEPRESSION</h2>
        {% endif %}
    </div>
    </center>
    </div>
<a href="{{ url_for('home')}}">Go back</a>
</body>
</html>
```

6.2 SERVER-SIDE CODING

Flask.ipynb

```
from flask import Flask,render_template,url_for,request
import pandas as pd
import joblib
# load the model from disk
clf = joblib.load("lr.pkl")
cv = joblib.load("lr_cv.pkl")
app = Flask(__name__)
@app.route('/')
def home():
    return render_template('home.html')
@app.route('/predict',methods=['POST'])
def predict():
    if request.method == 'POST':
        message = request.form['message']
        data = [message]
        vect = cv.transform(data).toarray()
        my_prediction = clf.predict(vect)
    return render_template('result.html',prediction = my_prediction)
if __name__ == '__main__':
    app.run(debug=False)
```

CHAPTER 7

SYSTEM TESTING

7.1 UNIT TESTING

The various modules of the system were developed and tested individually after the development of each unit. Each form was designed and each api in the flask app was tested after the integration with the UI. Each of the models was trained separately.

7.2 INTEGRATION TESTING

After the development and the testing phase of each of the modules has been completed, all the units were integrated into a single module. After the training and the testing of each mode in the unit testing phase, the pth and the pickle files were used in the flask app file to integrate all the functionalities. The flask app was developed and all the functionalities were tested.

7.3 TEST CASES & REPORTS

TEST CASE ID	TEST CASE / ACTION TO BE PERFORMED	EXPECTED RESULT	ACTUAL RESULT	PASS/ FAIL
1	Display the Home page by clicking on the Website Link	Display the Information and Features of the website	Display the features of the website	Pass
2	Enter the questionary and select the button predict.	Display the input text in questionary box	Display the input text in questionary box	Pass
3	Select the predict button.	Display the predicted result.	Display the predicted result.	Pass

CHAPTER 8
CONCLUSION

8.1 RESULTS AND DISCUSSION

Our machine learning project introduces a novel approach to enable automated prediction of depression based on a questionnaire. We have developed a deployment model using Flask, which allows users to enter textual data and receive predictions about their mental health. This system provides a convenient and accessible way for individuals to assess their risk of depression. Our project aims to contribute to the development of more effective and scalable methods for early detection and prevention of depression.

8.2 CONCLUSION AND FUTURE ENHANCEMENTS

In conclusion, the analytical process for this project involved several steps, including data cleaning and processing, missing value handling, exploratory analysis, and model building and evaluation. The highest accuracy was achieved on a public test set using a high-accuracy scoring algorithm, which was then used in the application to detect depression in individuals. To enhance this system, future work could involve implementing it in multiple programming languages, deploying it on a cloud platform for better accessibility, incorporating more data to improve accuracy, and adding voice input functionality. These enhancements can provide an even more robust and effective tool for detecting depression in individuals.

APPENDICES

A.1 SAMPLE SCREENS

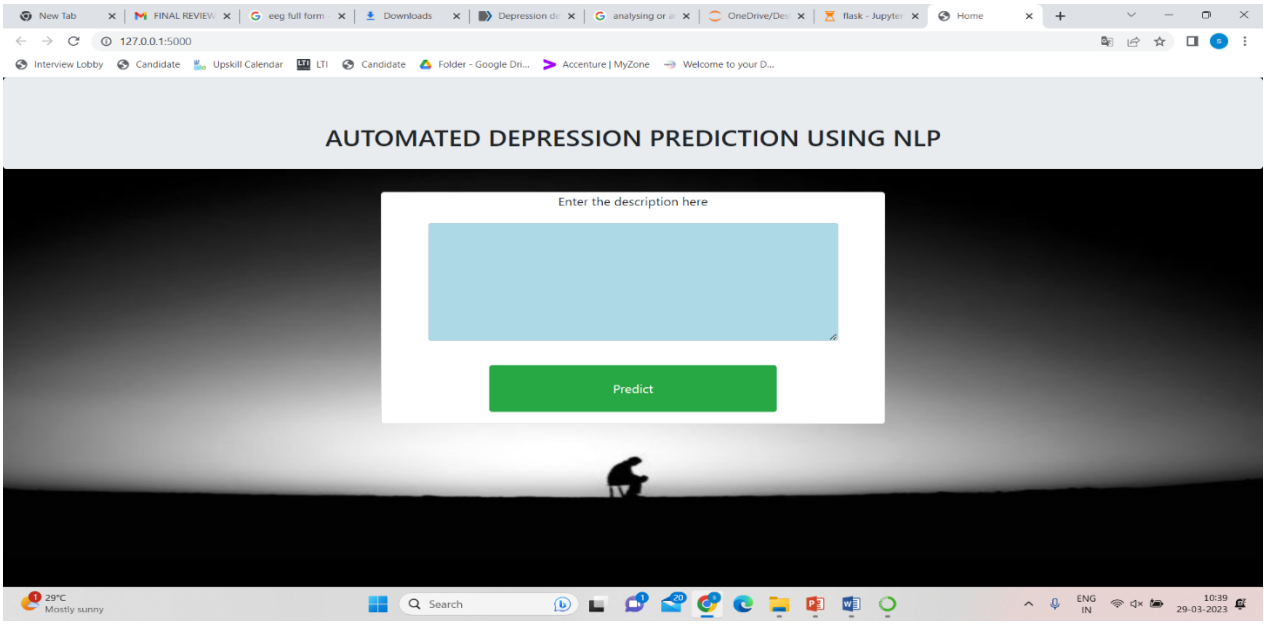

Fig.A.1.1 Website

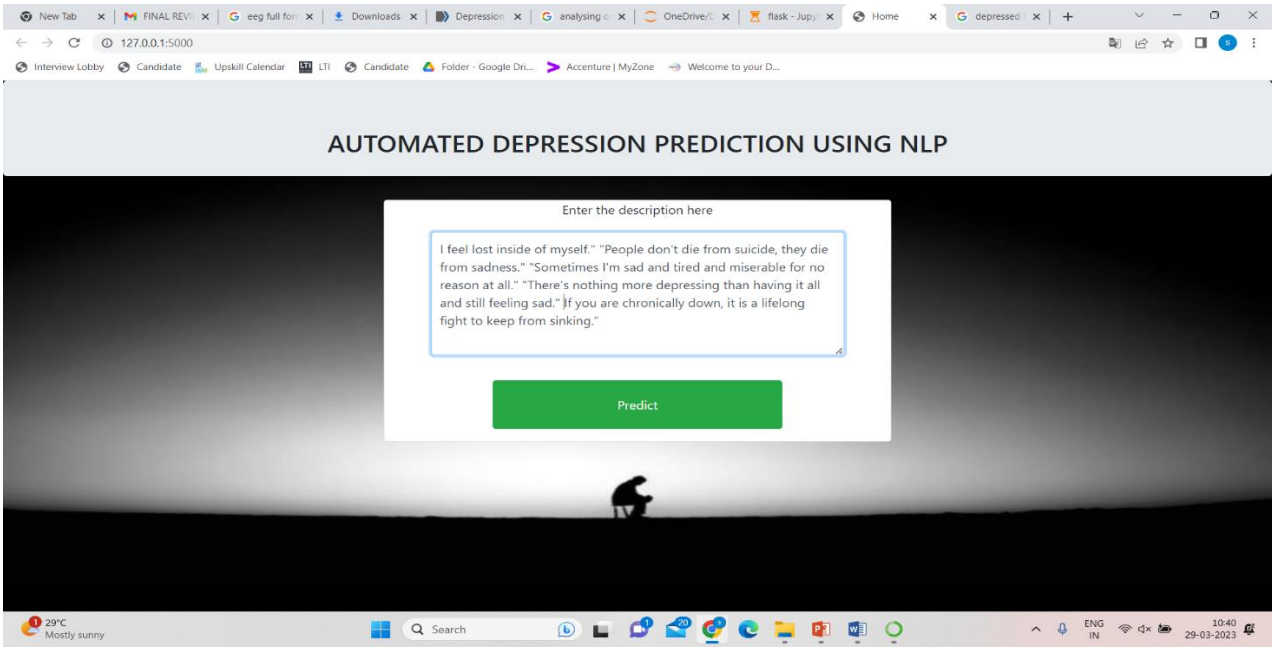

Fig.A.1.2 Input (for depressed person)

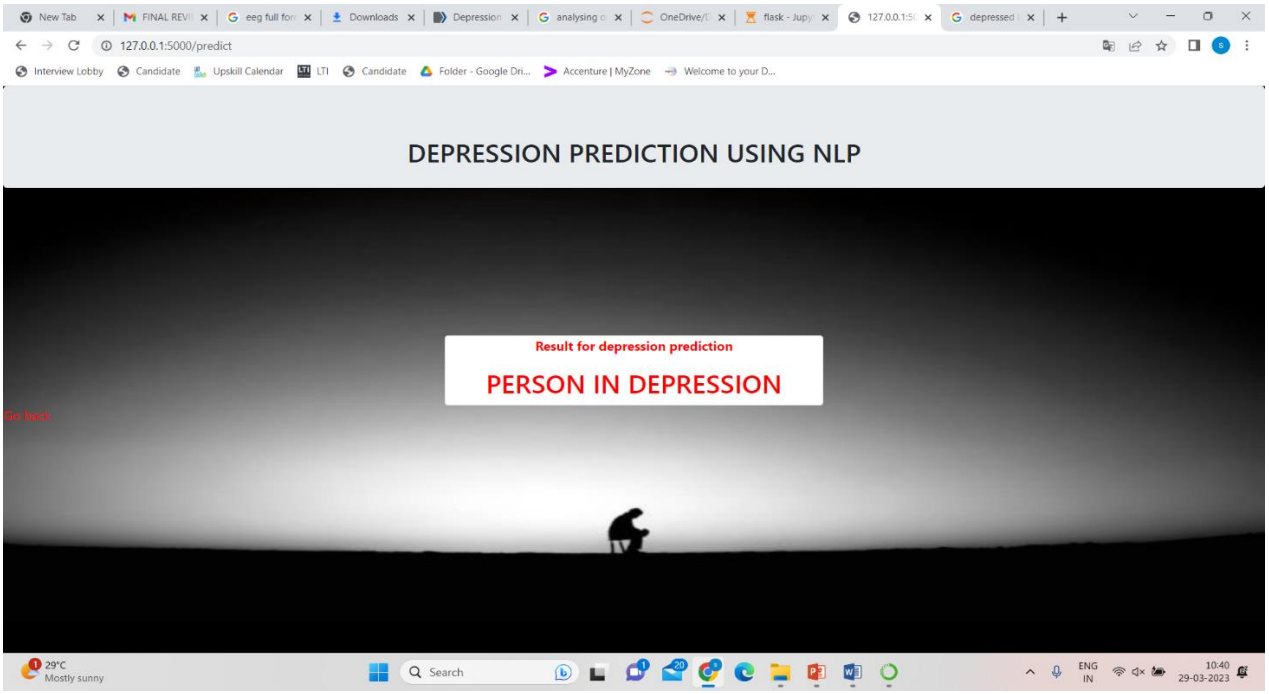

Fig.A.1.3 Predicted output (for depressed person)

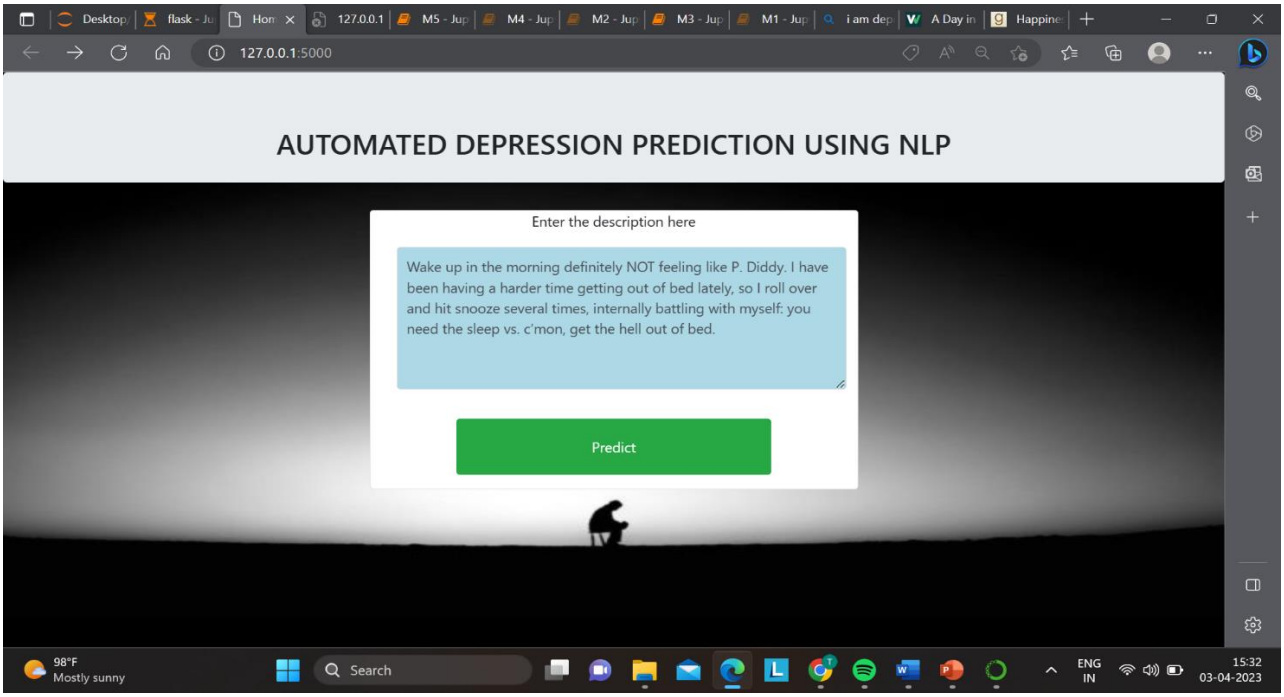

Fig.A.1.4 Input (for not depressed person)

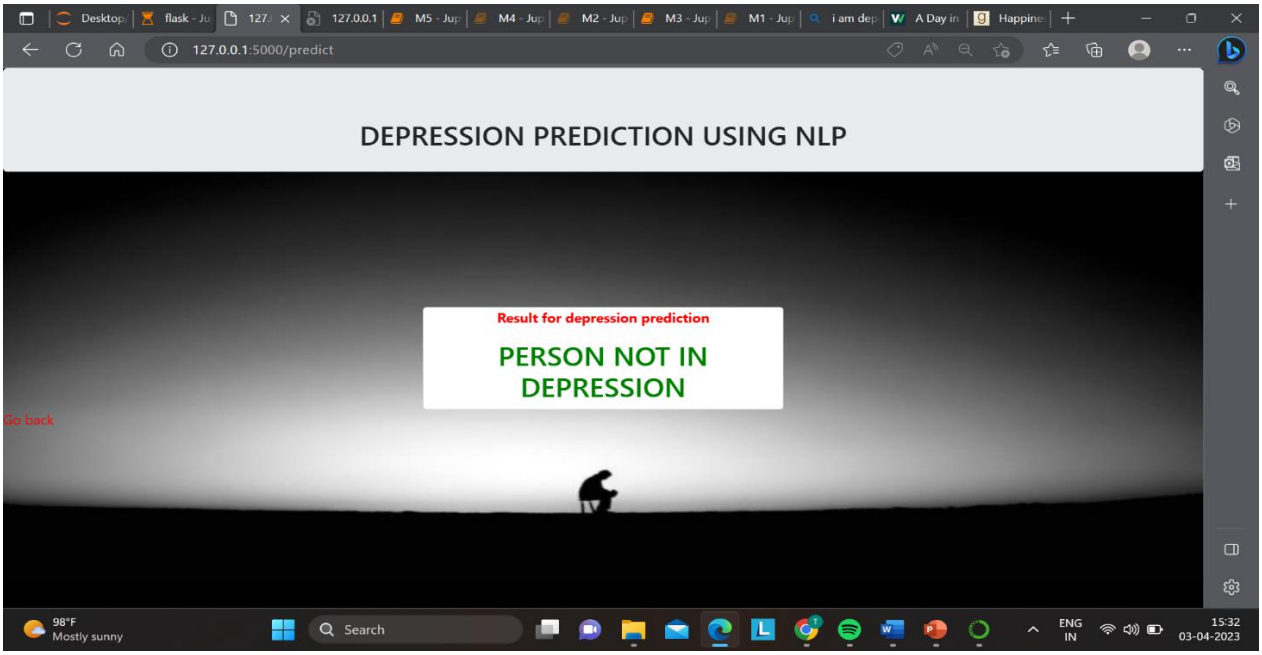

Fig.A.1.5 Predicted output (for not depressed person)

REFERENCES

[1] Meera Sharma, Sonok Mahapatra, and Adeethyia Shankar, “Predicting the Utilization of Mental Health Treatment with Various Machine Learning Algorithms,” WSEAS Transactions on Computers, vol.19,E-ISSN: 2224-287,2019.

[2] Usman Ahmed,Jerry Chun-Wei Lin, and Gautam Srivastava.“ Fuzzy Explainable Attention-based Deep Active Learning on Mental-Health Data,” IEEE International Conference, pp.6654–4407, 2019.

[3] Konda Vaishnavi, Nikhitha Kamath, Ashwath Rao, and N V Subba Reddy “Predicting Mental Health Illness using Machine Learning Algorithm,” Journal of Physics: Conference Series, 2021.

[4] Theodoros Iliou, Georgia Konstantopoulou, Mandani Ntekouli And Christina Lymperopoulou, “ILIOU Machine Learning Preprocessing Method for Depression Type Prediction” , pp.257-263, 2019.

[5] M. Kamal et al., "Predicting Mental Illness using Social Media Posts and Comments", vol 11, 10.14569/IJACSA.2020.0111271,2021.

[6] Srikanth Tadisettyet all., “Anonymous Prediction of Mental Illness in Social Media” Srikanth Tadisetty”, IEEE 11 Computing and Communication Workshop and Conference (CCWC), 2021.

[7] B Anishfathim, B Sreenithi,S Trisha,J Swathi and M Sindhu Priya, “The Impact of Mental Health due to Covid 19 – A Mental Health Detector Using Machine Learning,” Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), vol.147, 113193, Feb 2022.

[8] Vaibhav Jain, Dhruv Chandel, Piyush Garg, Dinesh Kumar Vishwakarma, “Depression and Impaired Mental Health Analysis from Social Media Platforms

using Predictive Modelling Techniques" Fourth International Conference on I-SMAC,2020.

[9] Reya Pillai R, Suchitra Saravanan, Gopal Krishna Shyam, "The BMI and Mental Illness Nexus: A Machine Learning Approach", International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2020.

[10] Piyush Kumar et al,"A Machine Learning Implementation for Mental Health Care" 11th International Conference on Cloud Computing, Data Science & Engineering ,2021.