Precision Medicine: Machine Learning in Depression Diagnosis

[1]Surendra Bandi, [2]Syed Zoya Mehak, [3]V. Vinod Chandra, [4]M. Soumya [5] P. Sai Samraat

[1]Professor, [2]Student, [3]Student, [4] Student,[5]Student,

[1]Computer Science & Engineering,

[1]Hyderabad Institute of Technology & Management, Hyderabad, India

Abstract: Depression is a prevalent condition that affects about 300 million individuals globally. Anyone can experience depression at any time. It could happen just once or again. It is not the same as typical mood swings or transient melancholy brought on by day-to-day issues. Future developments in technology are transforming the healthcare sector. The discovery of the causes of depression may result in novel studies and therapeutic approaches. as depression is quickly rising to the top of the global community health agenda. We suggested a method that uses machine learning (ML) algorithms to analyse user input data and utilise the results to predict whether or not the user is depressed. The proposed model was adopted to improve the classification performance, and the result showed classification accuracy of 99.74%.

Keywords - **Firebase, SVM, Random Forest, Secure Hash Algorithm**

1. INTRODUCTION

The prevalent and debilitating mental health issue of depression affects millions of people worldwide. It has an impact on an individual's ideas, feelings, actions, and general quality of life in ways that transcend beyond transitory mourning. The World Health Organisation (WHO) states that depression is the leading cause of disability globally, emphasising the critical need for effective protocols for identification, treatment, and assistance.

1.1 WHAT IS DEPRESSION? :

Regardless of a nation's status—developed or developing—one of the biggest issues facing the entire planet is healthcare. Smart, secure, and efficient healthcare systems are being developed as a top global priority to improve people's quality of life. Early research on human behaviour has drawn scientists from a variety of disciplines to the study of psychology and neuroscience.

This also holds true for the expanding fields of computer science and machine learning research. Finding a patient's mental health problems is a persistent problem for medical professionals and institutions, and it is not a recent development, particularly when it comes to younger patients. Recent developments in deep learning and machine learning have demonstrated the technology's ability to both diagnose psychiatric illnesses in people and determine how those disorders affect their way of life.

1.2. SYMPTOMS & PATTERNS:

During a depressive episode, a person experiences a depressed mood (feeling sad, irritable, empty). They may feel a loss of pleasure or interest in activities. A depressive episode is different from regular mood fluctuations. They last most of the day, nearly every day, for at least two weeks. Other symptoms are also present, which may include:

A. *Poor Concentration:*

Depression typically causes impaired concentration, which makes it harder to focus, make decisions, and retain memory. Although this cognitive impairment might affect daily functioning, it can be treated with the right therapy or medicine, leading to an improvement in overall mental clarity and functioning in depressed people.

B. *Feelings of Excessive Guilt or Low Self-Worth :*

People who are affected by depression often experience emotional discomfort because of their excessive guilt and low self-worth.

C. *Hopelessness About The Future:*

When depressed, it can be hard to see bright futures or opportunities, which often leads to a deep sense of hopelessness about the future.

D. *Thoughts About Dying or Suicide:*

People who are depressed may have terrible thoughts about ending their lives or committing suicide, which is an indication of the severe emotional pain and suffering they may be experiencing.

E. *Disrupted Sleep:*

Depression frequently causes sleep habits to be disrupted, which can result in excessive sleeping or insomnia, which can exacerbate the illness and negatively impact overall health.

F. *Changes In Appetite or Weight:*

Variations in appetite and weight are often associated with depression, which can lead to overeating or appetite loss and negatively impact an individual's physical and mental health.

G. *Feeling Very Tired or Low in Energy:*

Even everyday tasks might seem extremely tough and demanding while dealing with depression, which often results in excessive weariness and low energy.

Depression can lead to problems in the community, at home, at work, and in school, among other areas of life. Depressive episodes can be classified as mild, moderate, or severe based on the severity and quantity of symptoms, as well as how they affect the person's ability to function. Different patterns of depressive episodes exist, such as:

The three types of depression are single episode, which refers to an individual's first and only episode; recurrent, which refers to an individual who has experienced at least two episodes of depression; and bipolar, which alternates between manic episodes and depressive episodes. Manic symptoms include euphoria or irritability, increased activity or energy, and other symptoms like increased talkativeness, racing thoughts, increased self-esteem, decreased need for sleep, distractibility, and impulsive reckless behaviour.

1.3 FACTORS CAUSING DEPRESSION:

A complex interplay of social, psychological, and biological variables causes depression. Depression is more prone to occur in people who have experienced catastrophic circumstances, such as unemployment or bereavement. Depression can exacerbate an individual's current situation and the depression itself by increasing stress and dysfunction.
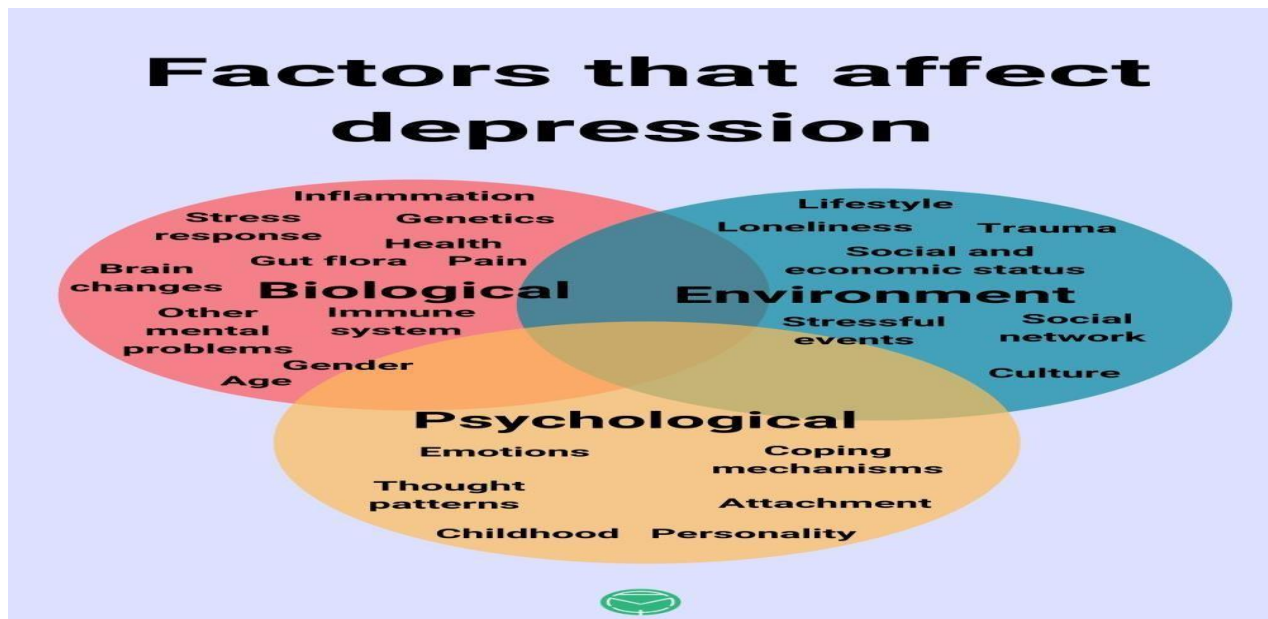

Figure 1: Factors That Affect Depression.

Depression and physical health are related to and have an impact on one another. Depression is impacted by numerous factors that are known to raise the risk of conditions like diabetes, cancer, heart disease, and respiratory ailments (such inactivity or risky alcohol use).People with a variety of illnesses may also develop depression due to the difficulties in managing their condition.

The following is a list of some of the Basic Factors:

1. Brain chemistry:
Depression may result from abnormalities in the amounts of certain brain chemicals.Brain chemistry abnormalities, specifically dopamine and serotonin imbalances, are linked to depression and have an impact on mood regulation and emotional health.

2. Genetics:
You may be more prone to depression if you have a depressed ancestor.Heredity has an impact on the risk of depression. A family history of depression may raise vulnerability because of inherited genetic factors and predispositions.

3. Trauma :
Experienced as a young child: Certain experiences influence how your body responds to stress and terror.Early childhood abuse or neglect can cause emotional and psychological damage that increases an adult's risk of developing depression.

4. Medical issues:
You may be more susceptible if you have certain conditions, such as attention-deficit hyperactivity disorder (ADHD), chronic illness, sleeplessness, or chronic discomfort.There are various medical conditions that can impact depression. Pain, long-term medical conditions, and alterations in brain chemistry can all contribute to or exacerbate depressive symptoms.

5. Drug use:
Your risk may be impacted by past drug or alcohol abuse.Excessive drug use can alter neurotransmitters in the brain, creating a chemical imbalance that raises the risk of depression.

2. LITERATURE SURVEY

Adolescent depression and other mental health issues have detrimental impacts on the kid, family, and society at large. Finding the causes of this mental health condition is essential. It's critical to recognise the warning signs of mental illness, such as depression, in kids and teenagers in order to make an early diagnosis and prevent serious repercussions down the road. No studies have been conducted using machine learning (ML) techniques to identify depression in kids and teenagers between the ages of 4 and 17 using a carefully crafted high prediction dataset like Young Minds Matter (YMM). Therefore, our goals are to: 1) develop a model that can predict depression in children and adolescents between the ages of 4 and 17; 2) assess the performance of ML algorithms to identify the best performing one; and 3) link with the associated problems of family activities and socioeconomic challenges that lead to depression.
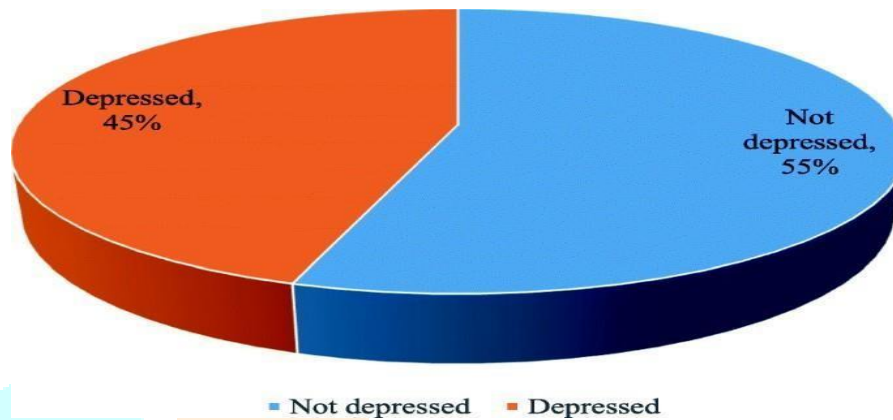

Figure 2: Globe population's depressive against non-depressive ratio

A global analysis of mental health uses a pie chart to illustrate the distribution of people's emotional well-being. According to the graph, 45% of individuals worldwide suffer from depression, even if 55% of the world's population does not now experience it.

These figures highlight the significant impact that depression has on a sizable portion of the global population.
A mood condition is what is known as depression. It is defined as emotions that interfere with a person's day-to-day tasks, such as grief, loss, or rage. Depression affects people differently. It could get in the way of your regular tasks, costing you time and output. It may also have an impact on certain long-term medical disorders and relationships.
https://www.researchgate.net/publication/331038027_Big_Data_and_Machine_Learning_Meet_the_Health_Sciences_Big_Data_Analytics_in_Mental_Health It is an approach that automatically learns the method and can provide a solution. Its learning process is related to human intelligence and used apparently in many intelligent environments like speech recognition service , self-driving vehicles and also provides recommendations (based on search history Google suggests what people want to search)[1]
https://pubmed.ncbi.nlm.nih.gov/30744717/ We employed a scoping review methodology to rapidly map the field of ML in mental health. Eight health and information technology research databases were searched for papers covering this domain. Articles were 12 RITW CSE assessed by two reviewers, and data were extracted on the article's mental health application, ML technique, data type, and study results.[2]
https://www.researchgate.net/publication/350086578_Persuasive_Technology_for_Mental_Health_One_Step_Closer_to_Mental_Health_Care_Equality The SPSRS application can address the three availability issues in mental health interventions: location-based availability (direct physical access to treatment), time-based availability (inability to receive needed treatment due to time of day), and cost-based availability (inability to access services beyond the recommended minimum weekly hours).[3]
https://pubmed.ncbi.nlm.nih.gov/26009311/ The study aimed at assessing the presence of anxiety and depression in cancer inpatients receiving palliative care at an oncology department using the Hospital Anxiety and Depression Scale (HADS) and determining whether anxiety and depression contribute to a lower quality of life controlled for pain and illness severity. Method: This cross-sectional study comprised 225 advanced cancer inpatients (a mean age of 65.1 years).[4]

3. EXISTING SYSTEM

Among all disorders, "depression" is thought to be the most complicated and dangerous psychological issue. It has a detrimental effect and is the main cause of illness burden. For this reason, a large number of scientists and medical professionals focused their research efforts on studying depression.

The World Health Organisation predicts that by 2030, depression would rank among the main causes of illness and mortality. Here, KNN algorithm is applied in the current systems. Even after receiving effective treatment, depression's effects persist in lowering performance and generally depressing a person's quality of life.

Support Vector Machine (SVM) is a robust and efficient approach for classification utilised in a multitude of practical contexts. SVM's efficiency and performance are mostly determined by its parameters. Compared to alternative classification techniques, SVM exhibits a greater rate of acknowledgement. Vapnik presented the SVM theory at Bell Laboratories. Even though SVMs at first were created as binary classifiers; nevertheless, other extensions have been suggested, such as manage the multiclass categorization situation. In order to oversee the division of the several classes, in these Extensions, extra constraints, and parameters are used on the optimisation problem.

4. PROPOSED SYSTEM

This article offers a comprehensive methodological framework for processing and analysing the heterogeneous data in order to better understand the relationship between depression and quality of life-related characteristics. It does this by utilising machine learning techniques. The experimental study is thus primarily split into two halves. A procedure for data consolidation is provided in the first section. Data relationships are established, and the Secure Hash Algorithm concept is used to uniquely identify each relationship in the data.

Finding and indexing the actual elements in the data is done by hashing. In the second section, a model that combined supervised and unsupervised machine learning methods was presented. The study hypothesis was formulated and validated with the aid of the consolidation approach. In order to further verify the effectiveness of the posterior probability multi-class Support Vector Machine, the classification problems were extracted from the clustered data and the self-organizing map yielded an 08 cluster solution.

The reliable and effective classification methods Random Forest and Support Vector Machine (SVM) are utilised in a variety of real-world applications. The parameters mostly determine the effectiveness and performance of SVM and Random Forest. Random forest and SVM have superior recognition rates when compared to other classification techniques. Although support vector machines (SVMs) were initially designed as binary classifiers, different adaptations, such as those for handling multiclass classification, have been proposed.

In these modifications, the optimisation issue is subjected to extra parameters and restrictions to control the division of the various classes. Multi-class problems are typically divided into many problems involving two classes using the standard techniques for applying SVMs to multi-class classification problems. Because of their exceptional learning efficiency, random forests and SVM have emerged as a hotbed of machine learning research.

5. METHODOLOGY

The following components are used in the construction of the Depression Detection using ML web application:

Key elements of the project:

Flask
Python
Numpy
Pandas

PYTHON:

Python is a popular general-purpose, interactive, object-oriented, high-level programming language. Python is a dynamically typed programming language that uses garbage collection. It was made by Guido van Rossum between 1985 and 1990. Similar to Perl, Python source code is likewise available under the GNU General Public License.

Python is a programming language that supports several paradigms, such as procedural, object-oriented, and functional programming. Significant indentation is used to promote code readability in the Python design philosophy.

Here are some advantages of Python:

1. Simple structure, few terms, and well-defined syntax make Python an easy language to learn. This facilitates the student's rapid language acquisition.

2. Wide-ranging standard library: The majority of Python's library is highly portable and compatible with Windows, Macintosh, and UNIX.

3. Interactive Mode: Python offers an interactive mode that facilitates the interactive testing and debugging of short code segments.

4. Portable: Python is compatible with a wide range of hardware systems and provides a uniform interface across them all.

5. Extendable − The Python interpreter can have low-level modules added to it. Programmers can enhance or modify their tools with these modules to make them more effective.

FLASK

Overview: The Flask API for Python enables the development of web applications. The creator of it is Armin Ronacher. Because Flask requires less basic code to construct a basic web application, it is easier to learn and provides a more explicit foundation than Django. A web-application framework, often known as a web framework, is a group of modules and libraries that enable programmers to create applications without having to write low-level code for thread management, protocols, and other features. Flask is built using the Jinja2 template engine and the WSGI (Web Server Gateway Interface) tools.

The following are some of the main advantages of utilising Flask:

a. It is a lightweight framework that makes development easier.
Permit the developer to experiment with their modules or architecture with flexibility.

b. It works well for modest projects.
Provides a quick debugger and integrated development server.

c. Support for secure cookies;

d. Applications that are easily scalable.

The online Server Gateway Interface (WSGI) is now the industry standard for developing online applications in Python. A standard interface for web servers and web applications is called Web Server Gateway Interface (WSGI).

It is a WSGI toolkit that implements utility functions, response objects, and requests.

Building a web framework on top of it is made possible by this. Werkzeug forms one of the foundations of the Flask framework.

A well-liked Python templating engine is called Jinja2. To render dynamic web pages, a web templating system combines a template with a specific data source.One common term for Flask is "micro framework." It seeks to maintain an application's core functionality basic and expandable.

Getting Started with Flask

A.Flask Installation :

Before we can run Flask, we first need to install the library. This can simply be done using the following command if you are using PIP.

pip install Flask

Note: It is required to have PIP installed before installing streamlit. This command will download and install Flask and its dependencies.

B. Building Flask

1. Create an App.py file: In your chosen Python IDE, create a new file called app.py.

2. Importing the libraries: The first coding step in the process is to import the required libraries.

import flask as Flas

C. Running Flask To confirm flask has been installed and imported correctly and that it can be run, go to the terminal and navigate to where your app.py file is located and run the following command:

python app.py

NUMPY:

An essential Python package for numerical and scientific computing is called NumPy (Numerical Python). Large, multidimensional arrays and matrices are supported, and a variety of mathematical operations can be performed on these arrays. For many scientific and data-related Python jobs, NumPy is a prerequisite. NumPy has various features including these important ones:

1. A powerful N-dimensional array object

2. Sophisticated (broadcasting) functions

3. Tools for integrating C/C++ and Fortran code

4. Useful linear algebra, Fourier transform, and random number capabilities

Advantages:

1. Efficiency: NumPy is substantially faster than Python lists for numerical calculations because of its highly efficient C and Fortran implementation. For manipulating arrays, it makes use of contiguous blocks of memory and optimized methods.

2. Data Integration: NumPy integrates well with other Python libraries for data analysis and scientific computing, such as SciPy, pandas, scikit-learn, and matplotlib.

PANDAS:

Pandas is a Python library used for working with data sets.It has functions for analyzing, cleaning, exploring, and manipulating data.The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008. Pandas gives you answers about the data. Like:

Is there a correlation between two or more columns?

What is the average value?

Max value?

Min value?

Advantages:

1. Excellent representation of data: The Pandas library is the perfect tool for anyone who wants to get into data science or data analysis because of the different ways it can represent and organize data. This is a very important function that cannot be disregarded because one can't possibly analyze or read any data unless it is represented well enough.

2. Efficient handling of huge data: This is the front that Pandas excels in. Wes McKinney made this library for the sole purpose of being able to perform on large amounts of data faster and better than each and every other library in the world. This makes it extremely important in analyzing copious amounts of data

6. FLOW OF EXECUTION

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So for this, we use data preprocessing task.

Data visualization is a crucial aspect of machine learning that enables analysts to understand and make sense of data patterns, relationships, and trends. Through data visualization, insights and patterns in data can be easily interpreted and communicated to a wider audience, making it a critical component of machine learning. In this article, we will discuss the significance of data visualization in machine learning, its various types, and how it is used in the field.

The train-test split is used to estimate the performance of machine learning algorithms that are applicable for prediction-based Algorithms/Applications. This method is a fast and easy procedure to perform such that we can compare our own machine learning model results to machine results. By default, the Test set is split into 30 % of actual data and the training set is split into 70% of the actual data. We need to split a dataset into train and test sets to evaluate how well our machine learning model performs. The train set is used to fit the model, and the statistics of the train set are known. The second set is called the test data set, this set is solely used for predictions.

The process of training an ML model involves providing an ML algorithm (that is, the learning algorithm) with training data to learn from. The term ML model refers to the model artifact that is created by the training process. The training data must contain the correct answer, which is known as a target or target attribute. The learning algorithm finds patterns in the training data that map the input data attributes to the target (the answer that you want to predict), and it outputs an ML model that captures these patterns

7. RESULTS & DISCUSSION

For medical professionals and organisations providing healthcare, diagnosing mental health issues is a difficulty. Our research, which makes use of NHANES data, indicates a relationship between some quality-of-life-related variables and depressed symptoms. The first stage in organising variables related to quality of life is clustering in order to more thoroughly examine the varied data associated with mental health issues. The data in the classes is readjusted and the clusters are reorganised by SOM to provide a balanced collection of groups. As a result, results have demonstrated that the suggested model may accurately predict the variables responsible with depression.
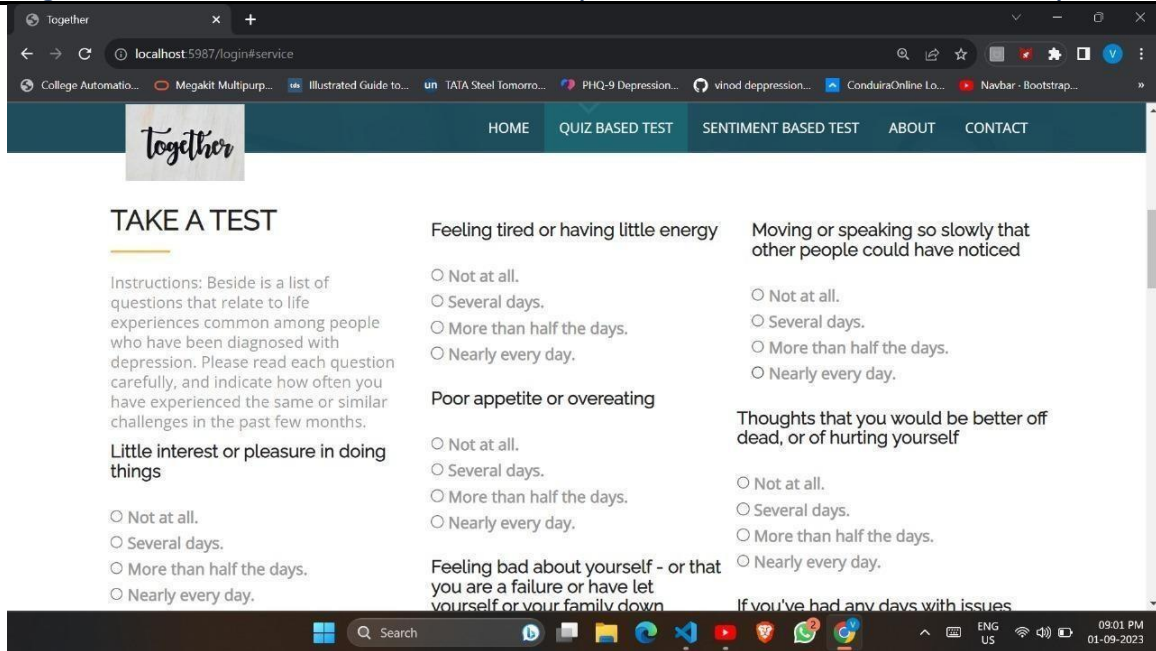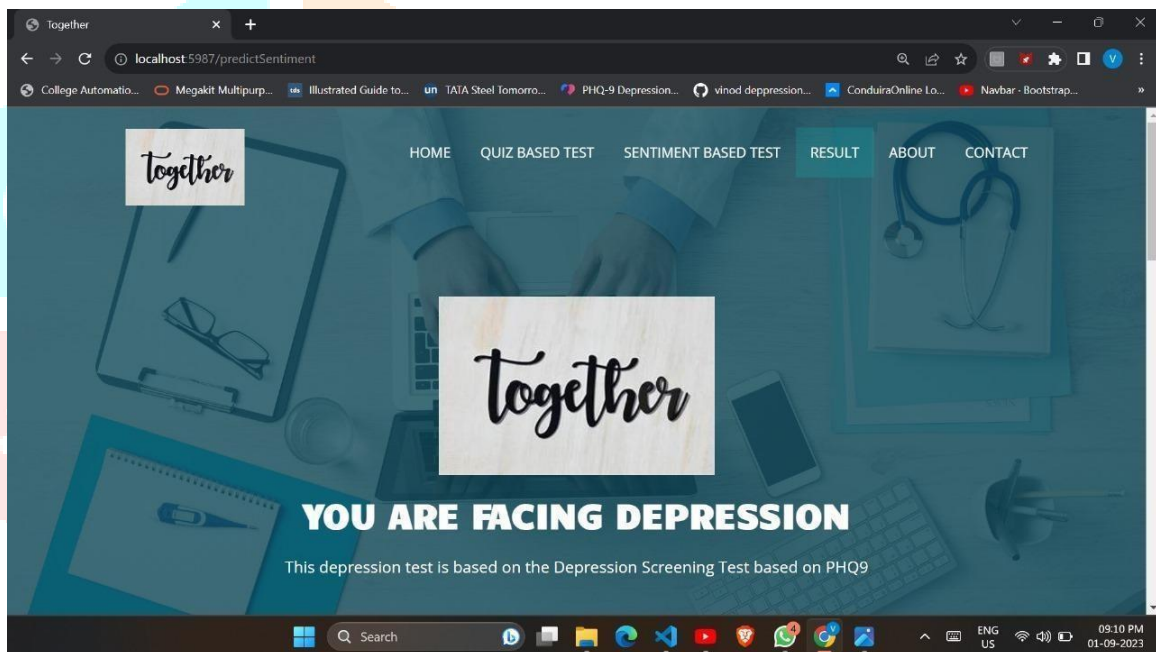
Figure 3: PHQ9 Questionnaires
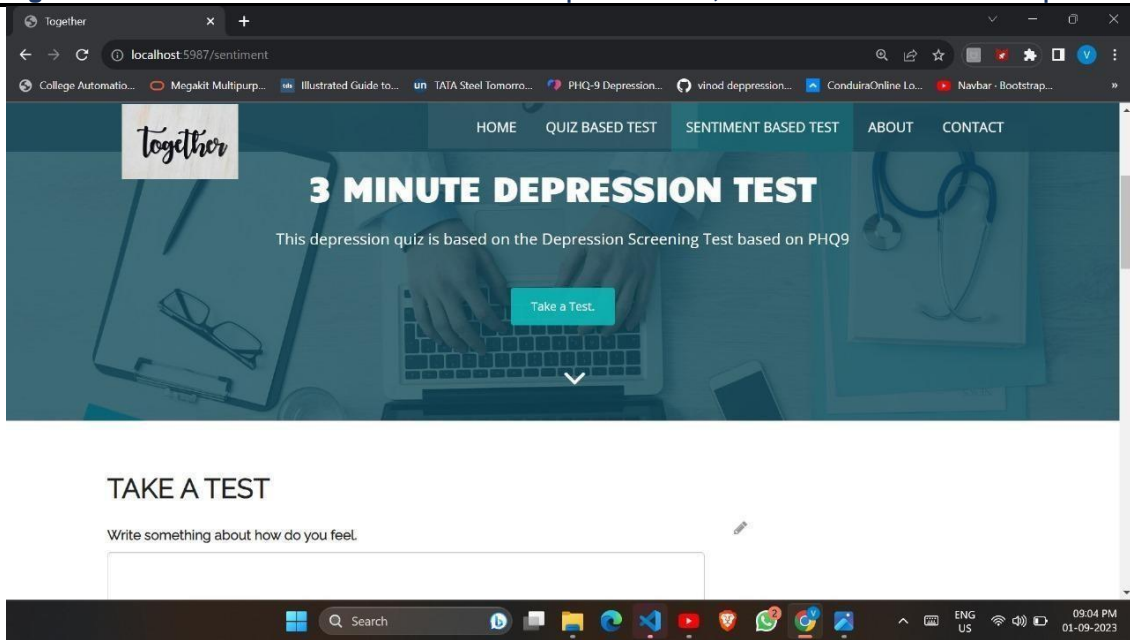

Figure 4: Results Page

Figure 5. Sentiment Test

8. REFERENCES

[1] I. C. Passos, P. Ballester, J. V. Pinto, B. Mwangi, and F. Kapczinski, "Big data and machine learning meet the health sciences," in Personalized Psychiatry, vol. 81, Cham, Switzerland: Springer, 2019, pp. 1–13.

[2] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: A scoping review of methods and applications," Psychol. Med., vol. 49, no. 9, pp. 1426–1448, 2019, doi: 10.1017/S0033291719000151.

[3] T. Kolenik and M. Gams, "Persuasive technology for mental health: One step closer to (Mental health care) equality?," IEEE Technol. Soc. Mag., vol. 40, no. 1, pp. 80–86, Mar. 2021, doi: 10.1109/MTS.2021.3056288.

[4] R. Bužgová, D. Jarošová, and E. Hajnová, "Assessing anxiety and depression with respect to the quality of life in cancer inpatients receiving palliative care," Eur. J. Oncol. Nursing, vol. 19, no. 6, pp. 667–672, 2015, doi: 10.1016/j.ejon.2015.04.006.

[5] L. Castelli, R. Torta, A. Mussa, P. Caldera, and L. Binaschi, "Fast screening of depression in cancer patients: The effectiveness of the HADS," Eur. J. Cancer Care, vol. 20, no. 4, pp. 528–533, 2010, doi: 10.1111/j.1365-2354.2010.01217.