STRESSOSPHERE

<u>A</u>

Project Report

Submitted in partial fulfillment for the award of the degree

<u>of</u>

Bachelor of Engineering in Information Technology

Submitted to



DEPARTMENT OF INFORMATION TECHNOLOGY UNIVERSITY INSTITUTE OF TECHNOLOGY

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL (M.P)

Submitted by

Anshita Verma(0101IT161010)

Shubhi Pande(0101EX161054)

Under the Guidance of

Prof. Asmita A. Moghe

(HOD IT Dept.)

Date:8th September 2020

Chapter 1: INTRODUCTION

1.1 Introduction

StressOSphere is a Xampp Application, an application for anyone to check the level of stress they are going through. It also provides some remedies and mental healthcare solutions. In this busy world people do not have the time to take care of their health. This leads to stress, anxiety and clinical depression. Anyone can have stress irrespective of their age, field of work, gender or financial conditions. We have gathered data within a diverse age group and field of work. It is high time now that we give attention to mental health the way we take care of our physical health.

Machine learning algorithms are implemented using the Python programming language. Each user will answer the questions given on the quiz as part of the survey. The algorithm is designed to analyze the features and suggest the level of stress that the person is going through. The stress Artificial level is estimated by the Neural Network algorithm.

1.2 Objective

The purpose behind the project is to develop awareness among the working and non-working class about mental health. Machine learning and Artificial Intelligence have been growing in almost every field of the world. In psychology as well, these technologies can be put to great use. The application will help the test takers to understand their stress level and work on it. This test can provide great information about how COVID19 has affected the mental health of individuals. This system will help to analyze how people reacted to the lockdown conditions and what will be the scenario post COVID19. The results not just include the stress level but also what can be done to overcome the stress.

1.3 Motivation

The motivation for the StressOSphere comes from the lack of concern towards mental health. No matter the financial status of an individual, most people are suffering with stress, anxiety, depression or more. But these emotions are generally suppressed. This often results in panic attacks and clinical depression.

Presently, a huge population is suffering from stress and anxiety in this lockdown. There is also a great increase in the cases of domestic violence in India. People around the world have no clue about the future which will bring even more stress. With our test, we can try supporting people who are going through high levels of stress in this difficult period.

Our motive is to draw attention towards the importance of mental health and help people take care of themselves. It is also to make people aware about the advantages of ML and how to involve it in psychology.

Many companies have already started using ML and the results have proved to be successful till now. With the advent of ever-increasing data in every field, it seems, the future is ML. The large data needs to be organized and should be taken for use and ML technology has the power to efficiently use this data and provide intelligence to machines.

And so, we intend to make people understand the importance of data and make them work with machines.

1.4 Scope

The Software provides an online test which will yield results as "low stress level", "Mild stress level" and "high stress level". It can be implemented with the help of internet. The application currently can analyze stress levels based upon user's input data.

Chapter 2: LITERATURE SURVEY

2.1 Introduction

In the current mechanism, there is no awareness for mental healthcare. This not only increases the stress level across but also leads to complex mental illnesses.

2.2 Introduction to Machine Learning

Humans appear to be able to learn new concepts without needing to be programmed explicitly in any conventional sense. We give a precise methodology for studying this phenomenon from a computational viewpoint. It consists of choosing an appropriate information gathering mechanism, the learning protocol, and exploring the class of concepts that can be learnt using it in a reasonable (polynomial) number of steps. Although inherent algorithmic complexity appears to set serious limits to the range of concepts that can be learned, we show that there are some important nontrivial classes of propositional concepts that can be learnt in a realistic sense.

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational

statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. IN[8]

2.3 Introduction to Neural Network

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural networks can adapt to changing input; so, the network generates the best possible result without needing to redesign the output criteria.

Neural networks, in the world of finance, assist in the development of such processes as time-series forecasting, <u>algorithmic trading</u>, securities classification, credit risk modeling and constructing proprietary indicators and price <u>derivatives</u>. A neural network works like the human brain's neural network. A "neuron" in a neural network is a mathematical function that collects and classifies information according to a specific architecture. The network bears a strong resemblance to statistical methods such as curve fitting and regression analysis.

A neural network contains layers of interconnected nodes. Each node is a perceptron and is like a <u>multiple linear regression</u>. The perceptron feeds the signal produced by a multiple linear regression into an activation function that may be nonlinear. IN[4]

2.4 Artificial Neural Network (ANN)

The idea of ANNs is based on the belief that working of human brain by making the right connections, can be imitated using silicon and wires as living neurons and dendrites.

The human brain is composed of 86 billion nerve cells called neurons. They are connected to another thousand cells by Axons. Stimuli from external environment or inputs from sensory organs are accepted by dendrites. These inputs create electric impulses, which quickly travel through the neural

network. A neuron can then send the message to other neuron to handle the issue or does not send it forward.

ANNs are composed of multiple nodes, which imitate biological neurons of the human brain. The neurons are connected by links, and they interact with each other. The nodes can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. The output at each node is called its activation or node value.

Each link is associated with weight. ANNs are capable of learning, which takes place by altering weight values. IN[4]

The following illustration shows a simple ANN -

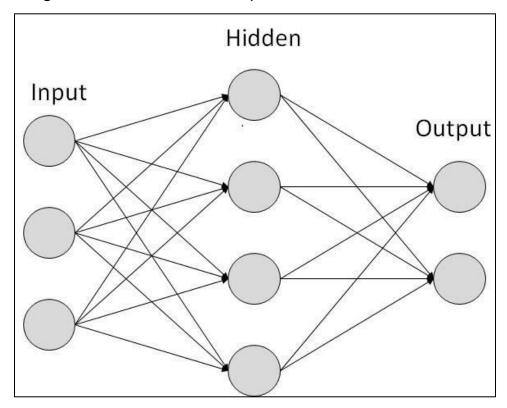


fig 2.4 ANN network architecture

2.4.1 Where do ANN come from?

Artificial neural networks are not a new concept. In fact, we didn't even always call them neural networks and they certainly don't look the same now as they did at their inception. Back during the 1960s we had what was called a perceptron. Perceptron was made of McCulloch-Pitts neurons. We even

had biased perceptron, and ultimately people started creating multilayer perceptron, which is synonymous with the general artificial neural network.

2.4 <u>Disadvantages of Current System</u>

- 1. There is no system available to check stress levels among people with an application.
- 2. It is not affordable for universities, companies or organizations to hire consultants and psychiatrists.
- 3. It is difficult to manage and pay attention to the mental health of employees and students.
- 4. No record of previous data for future analysis.

2.5 Advantages of Proposed System.

- 5. It could be used to analyze the mental health of employees in any organization.
- 6. Since it is based on ANN, all the available previous data would become useful to generate more accurate results.
- 7. It could also be used to test pressure or stress level among students.
- 8. Would be a more affordable option as it is an application only.

Chapter 3: PROPOSED SYSTEM

3.1Algorithm Used

1) ANN Algorithm

An artificial neural network is a supervised learning algorithm which means that we provide it with the input data containing the independent variables and the output data that contains the dependent variable.

In the beginning, the ANN makes some random predictions, these predictions are compared with the correct output and the error (the difference between the predicted values and the actual values) is calculated. The function that finds the difference between the actual value and the propagated values is called the cost function. The cost here refers to the error. Our objective is to minimize the cost function. Training a neural network basically refers to minimizing the cost function.

A neural network executes in two phases: Feed Forward phase and Back Propagation phase. IN[5]

a. Feed Forward

In the feed-forward phase of ANN, predictions are made based on the values in the input nodes and the weights. The weights of a neural network are basically the strings that we must adjust in order to be able to correctly predict our output.

The following are the steps that are executed during the feedforward phase of ANN:

Step 1: Calculate the dot product between inputs and weights Mathematically, the summation of dot product:

$$X.W=x1.w1 + x2.w2 + x3.w3 + b$$

Step 2: Pass the summation of dot products (X.W) through an activation function.

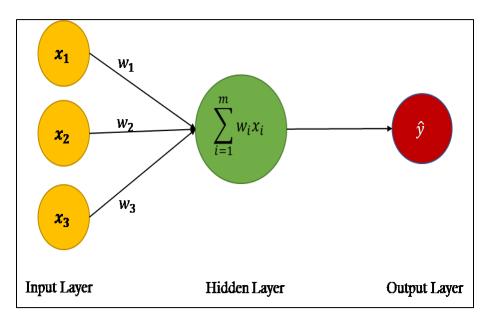


fig 3.1(a) Single layer Neural Network

b. Back Propagation

We start by letting the network make random output predictions. We then compare the predicted output of the neural network with the actual output. Next, we update the weights and the bias in such a manner that our predicted output comes closer to the actual output. In this phase, we train our algorithm. Steps involved in the backpropagation phase: -

Step 1: Calculate the cost

3.2 Analysis of data

1) Data Cleaning and Preprocessing

We need to handle missing values and categorical features before feeding the data into an artificial neural network algorithm, because the mathematics underlying most ANN models assumes that the data is numerical and contains no missing values. To reinforce this requirement, scikit-learn will return an error if you try to train a model using data that contains missing values or non-numeric values when working with models. We follow the following steps: -

c. Handle Missing Values

First, use the Pandas DataFrame method isnull() to return a DataFrame containing Boolean values:

- i. True if the original value is null
- ii. False if the original value isn't null

Then, use the Pandas DataFrame method sum() to calculate the number of null values in each column.

d. Investigation of Categorical Columns

To find out the number of columns that are of the object data type and figure out how we can make those values numeric.

2) Data Encoding

Since ANN algorithm requires data in numeric form, the categorical columns' data values are encoded. All the textual responses are given in numerical weights according to their significance. 'Yes' is taken to be 1, 'No' is taken to be 0 and a 'Sometimes' is taken to be 0.5. All 'NaN' cells are replaced with 0. The categorical data is converted into numeric data using encoder. IN[1]

The snapshot of the database after performing data preprocessing and data encoding is given in fig 3.2.

work status?	owers are	picture, is	, do you d	couldn't se	op of bad s	tuations a	ke you are	if you are	/el, how m	Stress Level
0	2	2	1	0	1	2	1	0	8	0
0	4	2	1	2	1	2	2	2	5	1
0	2	2	0	1	2	2	2	2	6	0
0	1	1	0	1	1	0	1	0	3	2
0	2	2	1	2	0	2	2	2	7	0
0	3	2	0	2	1	2	2	2	8	1
0	1	1	0	1	1	0	1	1	3	2
0	2	2	0	2	0	2	0	0	9	0
0	1	2	1	0	1	0	0	1	6	2
0	2	2	0	2	0	2	0	2	8	0
0	2	2	0	1	0	0	1	0	4	2
0	3	2	2	0	0	0	0	1	4	0
0	1	2	2	0	2	2	2	0	6	1
0	3	2	2	0	0	0	2	2	8	1
0	2	2	2	2	1	2	2	2	7	1
0	1	1	2	1	1	2	1	2	7	2
0	1	2	2	2	2	2	2	2	5	1
0	1	1	0	0	1	0	1	2	5	2
0	1	2	1	1	1	2	1	0	3	2
0	2	2	0	0	1	0	1	2	7	0
1	1	0	0	1	0	2	0	2	4	0
0	1	1	0	0	1	0	1	0	4	2
0	2	2	0	1	0	0	1	1	6	2
0	1	1	0	2	1	0	0	1	2	2
0	2	1	0	2	1	1	0	1	8	2
0	1	0	0	0	0	2	2	2	2	0
0	1	1	1	0	0	0	2	0	5	2
0	1	1	0	0	2	2	1	2	8	0

fig 3.2 Database after data preprocessing and encoding

Chapter 4: DESIGN

4.1 Backend Design

1) Low level database type used

The format of the database used in StressOSphere is **CSV** files. CSV is a simple file format used to store tabular data, such as spreadsheets or databases. Files in the CSV format can be imported to or exported from programs that store data in tables, such as Microsoft Excel or OpenOffice Calc.

CSV stands for "comma-separated values". Its data fields are most often separated, or delimited, by a comma.

A CSV file is a text file, so it can also be created and edited using text editor. More frequently, however a text file is created by exporting (File Menu -> Export) a spreadsheet or database in the program that created it.

To create a CSV file using a text editor, first choose your favorite text editor, such as Notepad or Vim and open a new file. Then enter the text data for the file, separating each value with a comma and each row with a line.

a. Database Description of StressOSphere

The database used in StressOSphere contains 19 attributes containing personal information about people and work life. In[2]

The attributes for which analysis is made is 'Stress level'. Stress level is the dependent variable. Out of so many columns about details of a person, there are eight attributes which influence the stress level with high intensity, these are questions related to checking their ability to cope up with stressful situations- "On a normal day, do you daydream a lot?, Do you feel like that you just couldn't seem to get going ahead in life?, Do you feel like you are stuck in a loop of bad situations and you can't get out of it?, Do negative thoughts and situations affect your daily functioning?, In a stressful situation,

do you feel like you are alone or you have no one to talk to?, Do you overreact to situations if you are under a lot of tension or stress, If you were to rate your stress level, how much would you rate it from 1-10?, According to you, what is the first reason that comes under your mind when you think about your stress? ". These eight attributes are independent variables. A snapshot of a database of candidate X with these attributes in .csv format are shown in fig 4.1.1.

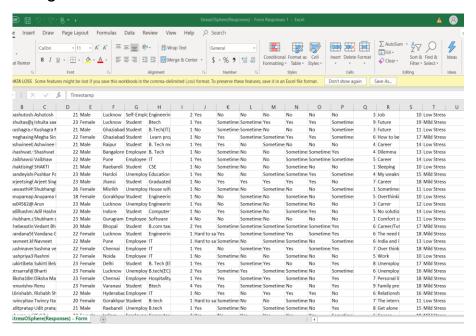


fig 4.1.1 CSV file of StressOSphere database

b. Database filtered for analysis and testing

The number of records present in the database is approximately five hundred, which is too large to take into analysis at a time. So, ten percent of the data is taken for analysis and testing. In[5]

The obtained dataset is then split into two categories of train dataset and test dataset. This is achieved using train_test_split function provided in the sk-learn package of python. The data in the train dataset is used by machines to learn and understand data. This gives intelligence to the machine to enable it to make predictions.

The test dataset contains the data for which we analyze stress levels of people and compare the predicted result with the actual value. This is used to analyze the error in the predicted values. By considering the percentage of values predicted incorrectly, we analyze the accuracy of the algorithm used.

2) Low level database

The low-level database csv file of StressOSphere is shown below in fig 6.3.2. It shows the data of each column in comma separated values form. This is the .txt file view of the database.



fig 4.1.2 StressOSphere Database

4.2 Frontend Design

The front end is designed using **Xampp.** It is a free and open-source cross-platform web server solution stack package developed by Apache Friends, consisting mainly of the Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages.

Xampp uses FTP protocol and makes the local server look like a remote server. The application StressOSphere runs on a local server using xampp.

The look of the final front end is provided in fig 4.2.

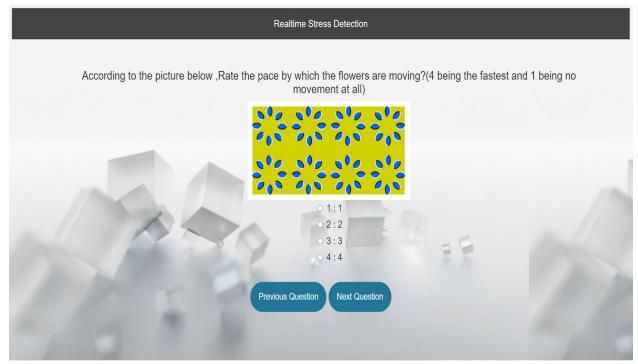


fig 4.2 Front End View

Chapter 5: RESEARCH OUTCOMES AND RESULT ANALYSIS

5.1 Result Analysis

From the data collected, the following figure 6.4.1 shows the proportions of people at different stress levels.

	Stress_Level
Low Stress	0.520243
Mild Stress	0.400810
High Stress	0.078947

fig 5.1.1 Stress Level Proportion

The survey for the StressOSphere project includes data of people from varied fields of jobs, places, educational background and covers different age groups from 15 years of age up to 70 years. So, to keenly analyze the levels of stress in different people on the, we used data visualization and deduced the following results.

a. Result analysis based on Gender

The comparison of stress levels with respect to gender shows the ratio of females is higher in High stress level whereas in ratio of male is more in Mild and Low stress level.

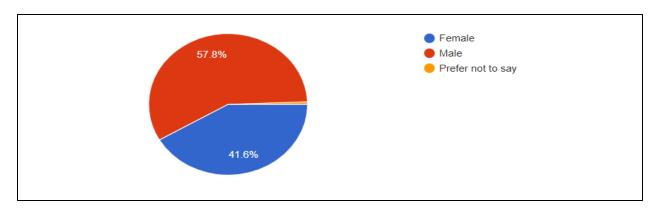


fig 5.1.2 Gender Proportion

Gender Stress_Level	Female	Male	Prefer not to say
High Stress	0.564103	0.435897	NaN
Low Stress	0.338521	0.657588	0.003891
Mild Stress	0.494949	0.500000	0.005051

fig 5.1.3 Stress Level Proportion by Gender

b. Result analysis based on Age and Gender

The statistics of stress level w.r.t gender are insufficient to deduce specific conclusions from the database. So, another parameter is taken along gender i.e., age and the data are divided into various age groups.

Fig 6.4.5 shows that people from lower age groups (15-25) have a higher ratio of High Stress. As we go to higher age groups the ratio of majority falls under Low Stress Group.

From this we can draw a conclusion, people of the new generation are facing more stress issues.

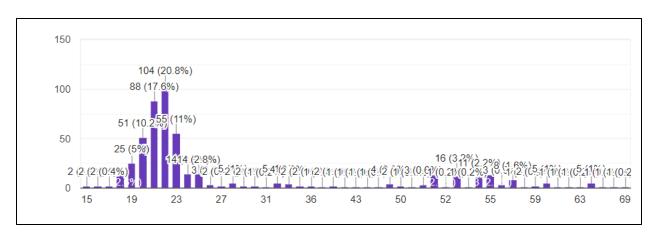


fig 5.1.4 Age Proportion

	Stress_Level	High Stress	Low Stress	Mild Stress
Agegroup	Gender			
15-25	Female	0.125749	0.341317	0.532934
	Male	0.067010	0.541237	0.391753
	Prefer not to say	NaN	0.500000	0.500000
25-35	Female	0.100000	0.500000	0.400000
	Male	NaN	0.437500	0.562500
35-45	Female	NaN	1.000000	NaN
	Male	NaN	0.833333	0.166667
45-55	Female	NaN	0.714286	0.285714
	Male	0.056604	0.735849	0.207547
55-70	Female	NaN	0.933333	0.066667
	Male	0.071429	0.785714	0.142857

fig 5.1.5 Stress level proportion w.r.t Age and Gender

c. Result analysis based on Work Status

Another parameter taken for comparison of stress levels is Work Status. Fig 6.4.6 shows the ratio of students in high and mild stress levels are higher. Also, people who are self-employed have a higher ratio in Low Stress.

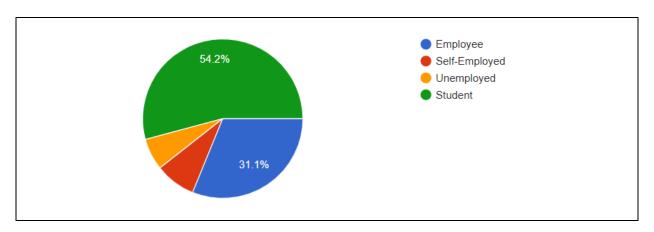


fig 5.1.6 Work Status Proportion

Work_Status Stress_Level	Employee	Self-Employed	Student	Unemployed
High Stress	0.333333	0.051282	0.615385	NaN
Low Stress	0.354086	0.116732	0.463035	0.066148
Mild Stress	0.257576	0.035354	0.631313	0.075758

fig 5.1.7 Stress level Proportion by work status

d. Result analysis based on Age and Work Status

The analysis with respect to age and work status shows people who are selfemployed and people from an older age group have a higher ratio in Low Stress than employees and students.

	Stress_Level	High Stress	Low Stress	Mild Stress
Agegroup	Work_Status			
15-25	Employee	0.147059	0.455882	0.397059
	Self-Employed	NaN	0.777778	0.222222
	Student	0.090566	0.441509	0.467925
	Unemployed	NaN	0.380952	0.619048
25-35	Employee	0.047619	0.476190	0.476190
	Student	NaN	NaN	1.000000
	Unemployed	NaN	0.500000	0.500000
35-45	Employee	NaN	1.000000	NaN
	Self-Employed	NaN	0.666667	0.333333
	Unemployed	NaN	1.000000	NaN
45-55	Employee	0.020408	0.734694	0.244898
	Self-Employed	0.117647	0.705882	0.176471
	Unemployed	NaN	1.000000	NaN
55-70	Employee	0.071429	0.785714	0.142857
	Self-Employed	NaN	0.900000	0.100000
	Unemployed	NaN	1.000000	NaN

fig 5.1.8 Stress Level proportion w.r.t age and work status

5.2 Result Outcome

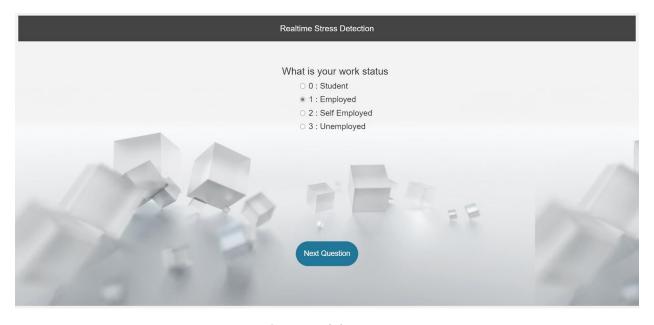


fig 5.2.1(a) Input

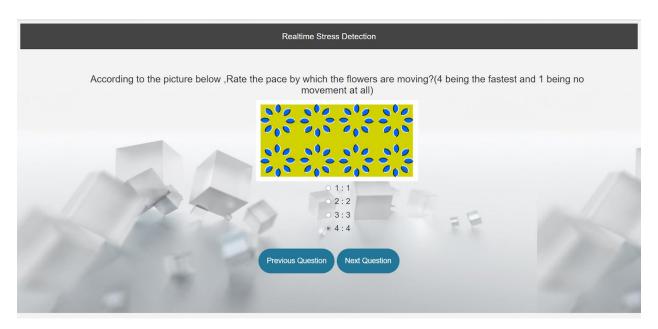


fig 5.2.1(b) Input

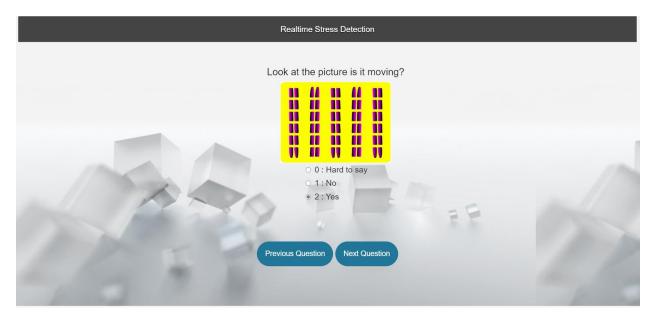


fig 5.2.1(c) Input

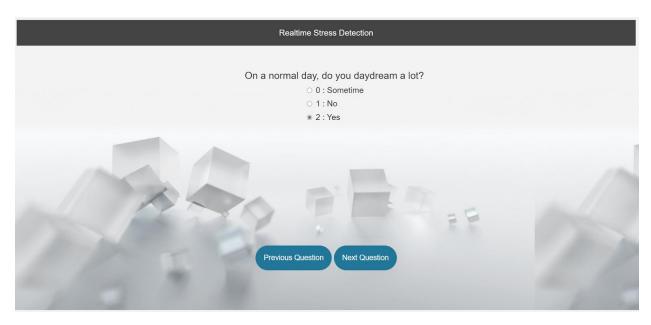


fig 5.2.1(d) Input

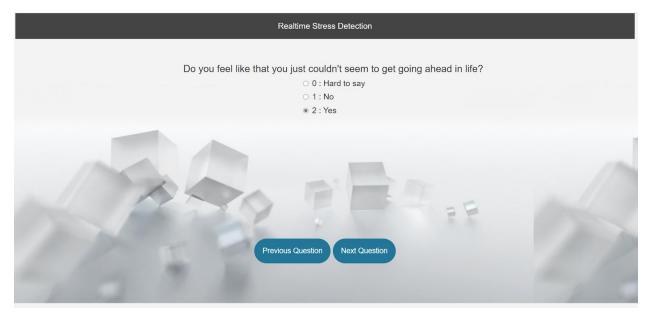


fig 5.2.1(e) Input

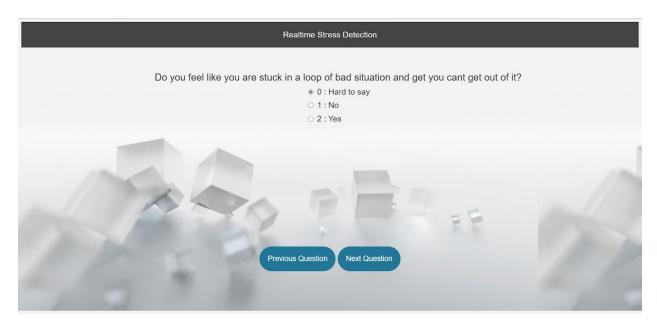


fig 5.2.1(f) Input

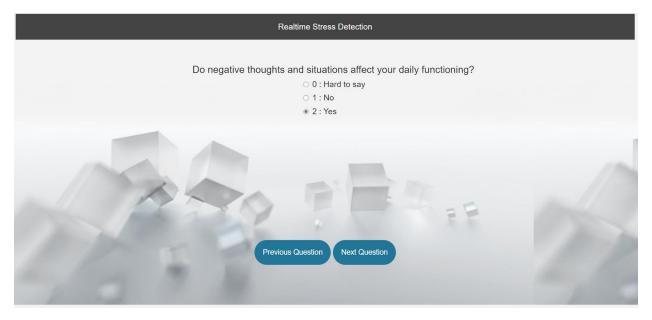


fig 5.2.1(g) Input

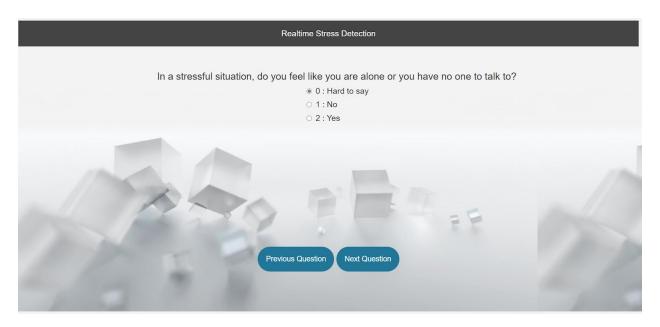


fig 5.2.1(h) Input

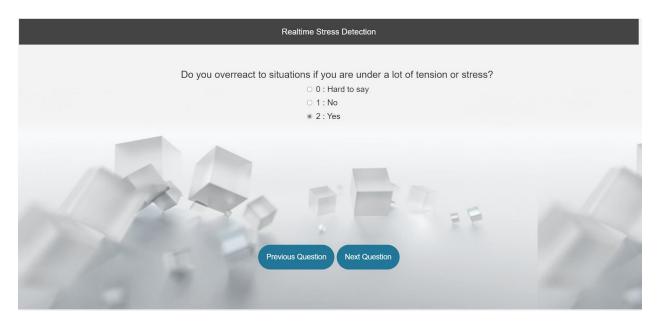


fig 5.2.1(i) Input

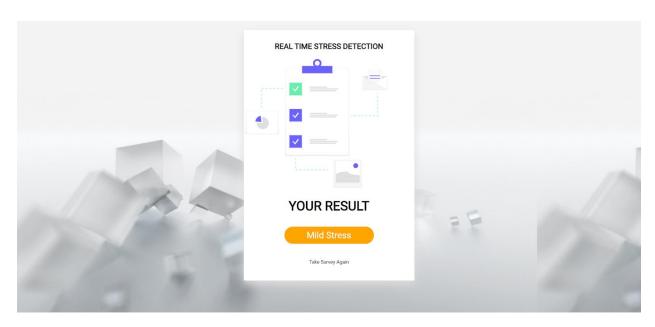


fig 5.2.1(j) Output

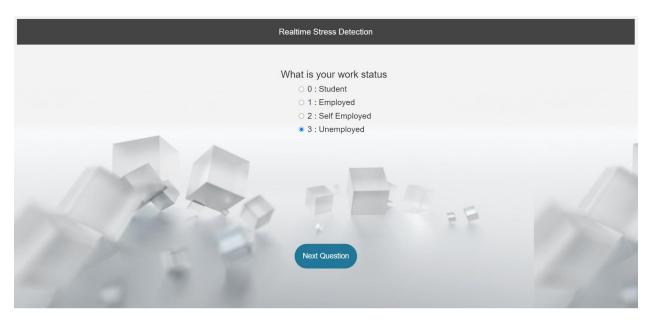


fig 5.2.2(a) Input

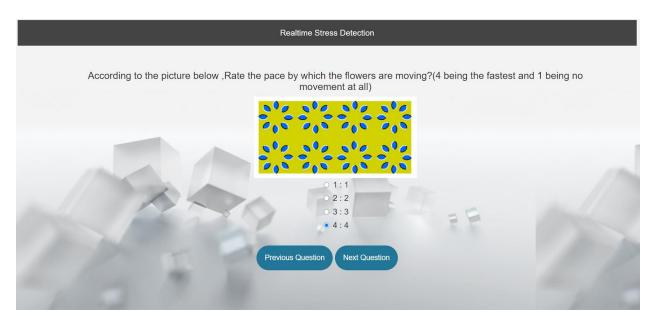


fig 5.2.2(b) Input

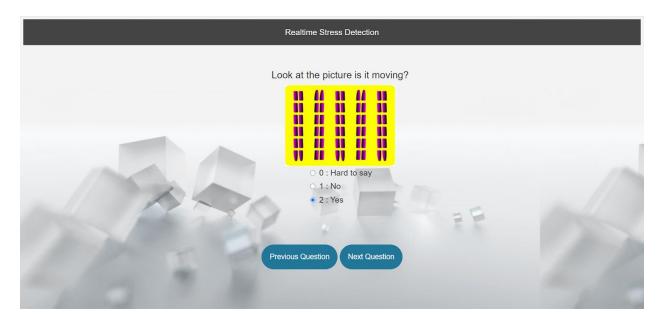


fig 5.2.2(c) Input

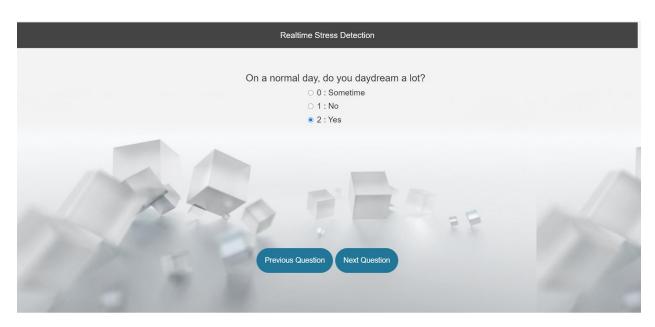


fig 5.2.2(d) Input

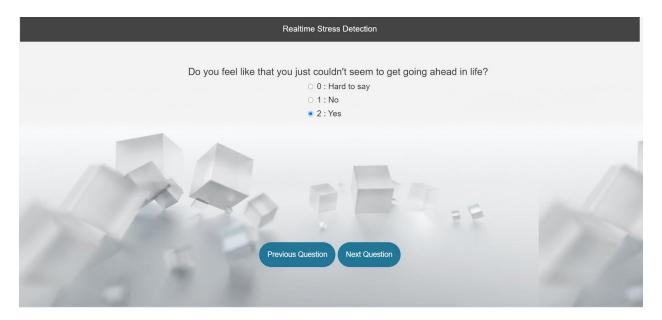


fig 5.2.2(e) Input

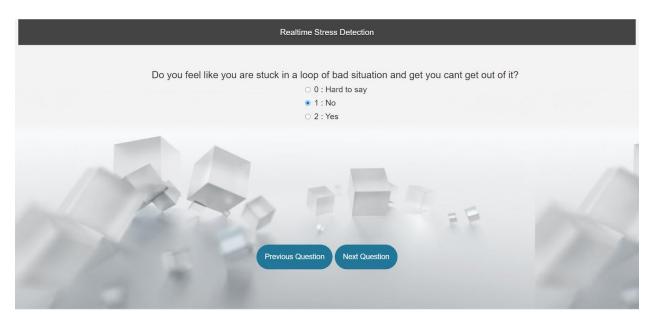


fig 5.2.2(f) Input

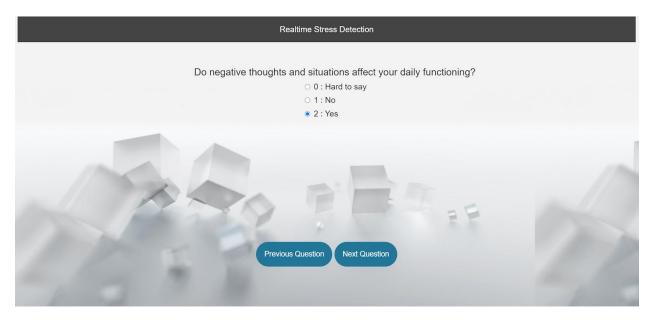


fig 5.2.2(g) Input

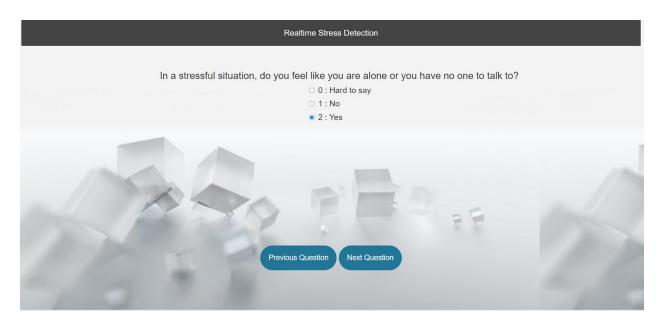


fig 5.2.2(h) Input

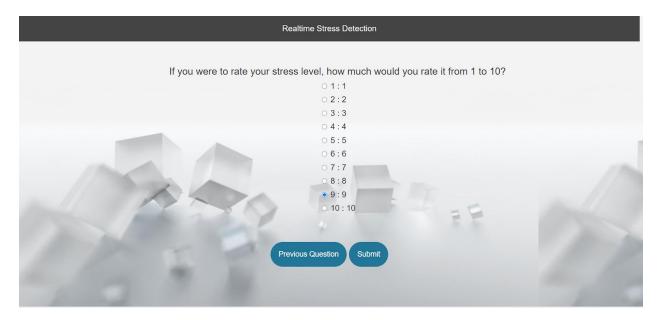


fig 5.2.2(i) Input

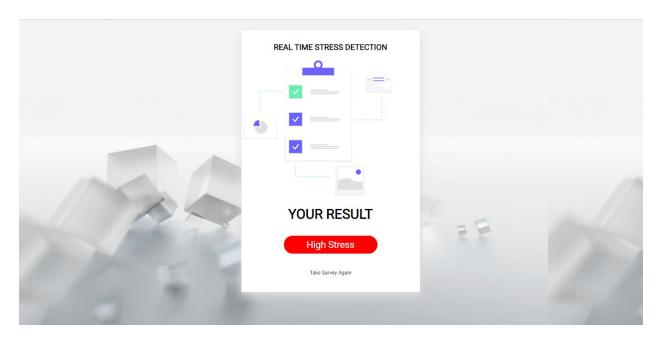


fig 5.2.2(j) Output

Chapter 6: CONCLUSION

The project is successfully implemented for prediction of Stress level with the given attributes. The use of Artificial Neural network algorithm enabled us in analyzing the data and giving results within fraction of seconds. The task of making a machine intelligent by training it with the supervised data provided to it and testing the results is achieved. The application can now assist humans with data analysis for and help in reducing stress in individuals.

The application can prove to be of great importance to doctors and individuals. The application will not only work according to them but the main purpose of using intelligent applications is to make humans work along with the machines. And the doctors will work to enhance the result of prediction to offer their patients and help more people by taking steps towards being healthy mentally.

Chapter 7: REFERENCES

- 1. Aditya Thotha and A.Dharun, Machine Learning Techniques for Stress Prediction in Working Employees, IEEE,2018.
- 2. Manvi Panchal, BA Honours, Psychology, IEHE University, MP, India.
- 3. https://www.stress.org/holmes-rahe-stress-inventory
- 4. https://www.coursera.org/
- 5. https://www.udemy.com/
- 6. https://onlinecourses.nptel.ac.in/noc19 cs35/course
- 7. www.wikipedia.com
- 8. https://data-flair.training/blogs/machine-learning-classification-algorithms/
- 9. www.stackoverflow.com