iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.43027

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Machine Learning Based Analysis on Behavioural Difference Between Depression and Anxiety Symptoms

Ayushi Saxena[1], Tanya Dhyani[2], Divyanshu[3]
[1, 2, 3]*Student, Department of Computer Science, Electronics and Communication , Electronics and Communication, Raj Kumar Goel Institute of Technology, Ghaziabad, India*

Abstract: ***Depression and anxiety are two different versions of mood disorders. It is a state in which people used to develop some deep side effects. Depression and Anxiety can develop some serious effects on individual in terms of both physical and mental emotions. It's main feature is sadness in extreme measures, and can often lead to suicides and mental traumas. Depression is a major illness that affect human minds in a worst sequence. According to a report by WHO (World Health Organisations) , dated June 2019-2020 , India is the most depressed country in the world, due to Covid situations and financial problems with 7% of its total population being traumatized of mental problems. And to take this problem as an illness, the first thing to do is analyzing the main difference . According to Anxiety and Depression Association of America (ADAA) anxiety and depression are the most common mental illness among people around United States. If you don't treat your anxiety disorder you will soon develop some extreme psychological disorders and later suffer with depression . Depression cause people to have sad mood, loss of interest , low esteem , bad taste , sleeping attacks etc. Our application mainly contains mental health problem quiz questionnaire to detect the early symptoms. But this analysis can be deceived easily if a patient tried to answer differently. Hence, we come here to provide one way method to treat and detect depression and anxiety detection using Machine Learning Algorithms. We obtain major questionnaire in the form of quizzes , and users fill it on basis of their mental emotions. This machine learning data is used to analyze the early detection of mental traumas on normal people. This algorithm takes readings of emotions from the answer of quizzes given by the specific user. Classification machine learning algorithm used K nearest neighbour, Naive Bayes theorem, Decision Tree Algorithm and Random Forest classifier have been used for the detection model, Word Clouds to differentiate the positive and negative words. After all the final results the user is immediately warned of their depression level, and they are urged to get professional help given by our website. The overall model not only achieves high accuracy due to its Machine Learning Approach model , but also inherits the scalability regarding one proper input size.***

Keywords: Depression detection, Anxiety detection, Machine Learning, Text classification, Supervised Learning , Word Cloud, K-Nearest Neighbors

I. INTRODUCTION

Machine learning algorithms is a subset of Artificial intelligence and it work accordingly . Machine learning actually mainly focuses on the development of programs that can access data from the real world technologies , and learn from its experiences with data formats . Now a days the early signs of depression detection is one of the most important field of psychology . Depression is typically a mental issue now a days . Recent researches also proposes that recent researches reveals that depression is also the main cause of disability and somatic diseases . The process of analyzing begins with deeply observations of data. The social media platforms like facebook , whatsapp , twitter holds lakhs of data every second . Every message or contents offer their users to express their feelings , emotions and sentiments of topic . The primary focus is to allow the computer system to learn automatically without any human intervention . Supervised machine learning algorithms use a set of labelled and set datasets to train themselves accordingly , and hence named supervised machine learning algorithm .The labelled dataset has to predict the future of data from the past . The observation of training dataset with known categories enables the learning algorithm to produce a function that makes predictions about the output variables that are needed. The machine learning algorithm can also compare its output with the correct output and translate its model accordingly. Depression is a mental illness that affects person moods , mental thoughts , and basically their day to day activities . The hardwork level and the focus of a person suffering from depression will decreased day by day , they will constantly be their own lonely life and nothing they do just happy and find joy in life without any reasons . Nowadays the early detection of depression and anxiety is typical illness in one's life. This is one of the most newly features of machine learning algorithms. Unsupervised learning data mainly occurs when the data is not fixed or not labelled. Unsupervised learning data helps the computers to describe a function that can identify the hidden pattern of any unlabeled dataset. This is because semi supervised machine learning algorithm uses both labeled and unlabeled data for training the model.

©IJRASET: All Rights are Reserved

Typically, the dataset contains a small amount of labeled data set and remaining large amount of data being unlabeled . Reinforcement machine learning algorithms depends on a learning path that interacts with its system by giving certain actions that are later work for error/reward method . This reward and error search rewards are the most advanced characteristics of reinforcement machine learning algorithm. This algorithm allows machines to significantly determine the ideal performance that can maximize or may be minimize its performance, even within a proper context. The simple reward system is necessary for the algorithms to analyze and learn the best delivers output; this is known as the reinforcement signal for a system machine learning . Machine learning algorithm enables analysis of massive number of quantities of data worldwide, with minimum to zero human interference . Combining machine learning with AI can make a machine learning model even more effective while processing large volumes of data processing. Classification algorithms and regression algorithms are the two types of supervised machine learning algorithm. The Classification algorithms are used when the output is restricted to a fixed set number of values. For example, for a classification algorithm that filters spam emails, classify positive and negative thoughts according to the quality of words..

II. LITERATURE REVIEW

A. *Existing System*

Authors of [3] planned that the behaviour of users in social networking is actively developing In this method, the computarization are successfully used in analyzing the posts and contents of social media. The results had analysis the differences in patterns between the two users, with around 70% of accuracy in differential classification. Authors of [4] had conducted critical appraisal of around 8000 of individuals. In this research, machine learning is used to process the scrapped data collected from Social Networking Sites users , Natural language processing, Support vector machine, Word Cloud of data, Data processing to detect depression in potentially and efficient way. The experiment is used to analyze the depression level of a person by observing and analyze the emotions by the normancy of the text. The experimental results show that the psychologists of the Italy are the most successful in analyzing depressed individuals with 90% and 91.9% proper accuracy. Authors of [6] had generated confusion matrix for the predictive analysis for using the High Alert Drug classification. This matrix is used for further prediction of mental health problems. The results delivered through the Random Forest (RF) classifier and Naïve Bayes based classifier have the highest accuracy rate of 89% to predict the exact results. Authors of [11] used LIWC Linguistic Inquiry and Word Count Tool for the classification tool of positive and negative effects to determine the differential analysis of depression and anxiety. Authors of [12] talk about the various advances in the natural language processing techniques and their characteristics. Many techniques such as Named Entity Recognition and Machine Reading have been analysed. Authors of [14] analyse and compare the various sources of data in internet from which data can be collected for natural language processing tools. The article by The Pharma Innovation of mental health problems talks about how depression is a most common issue in many people nowadays and it occurs in all age, caste and background. The three most common types of depression, which are major depression, dysthymia and bipolar disorder, minor depression. Major depression is mainly by feeling of sadness , heartbroken or nonexistive mood which lasts longer for two weeks and the victim can't sleep properly, can't eat or enjoy good activities. Dysthymia is a less form of maximum depression disorder in human beings. However, it is long lasting by preventing a person from feeling good. Authors of [18] focus on new dimensions of social media data to detect the early depression signs. Their results show that the decision tree algorithm has the highest accuracy in emotional process style.

B. *Proposed System*

The proposed system detects depression in the user through machine learning technique. The system analyses the text dataset collected from people who were clinically diagnosed with depression. This helps the machine learning model to detect the traces of depression in the user's submitted input text. For the input, the user submits a random piece of text written by them to the application. This means that the proposed system does not access to the sensitive social media content of the user. Furthermore, this input is voluntarily bestowed by the user, which means that there is no privacy breach. The user can access the system in the comfort of his privacy, that protects him from surrounding social stigma, As the fear of social stigma is eradicated from user, the accuracy of the input data that is fed into the proposed system also increases. The machine learning model trained in text classification algorithms analyze the input text and produces its current depression diagnostics. As depression is not a spontaneous occurrence, and needs at least a week to set in, we consider pieces of text from previous times and combine their results to pass a verdict. The conclusion can be very depressed, mildly depressed, or not depressed. The final output consists of a brief description of the conclusion, assisted with the graphical representation of depression levels from various times.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

The purpose of the graphical representation is to represent the depression patterns of the user. The detected pattern can be helpful for the user, to understand their emotional swings. This depression pattern from the graphical representation consists both the past depression readings, along with the present one.

III. BAYES THEOREM

Bayes theorem is a statement that doesnot allow exchange of the order of the events. In other words, if A and B are two events , the occurrence probability of the event A given B is not the same as the occurrence probability of the event B, given A. In the case of mental health psycologists, if we develop a test to determine to estimate if a person is normal or suffering from depression . There may be two cases which individually analyze the better accuracy . The probability estimation analyze that if he/she is positive it shows that the person is normal and if he/she shows negative it predicts the negative feedback of the mental health. By indicating that we psychologically.

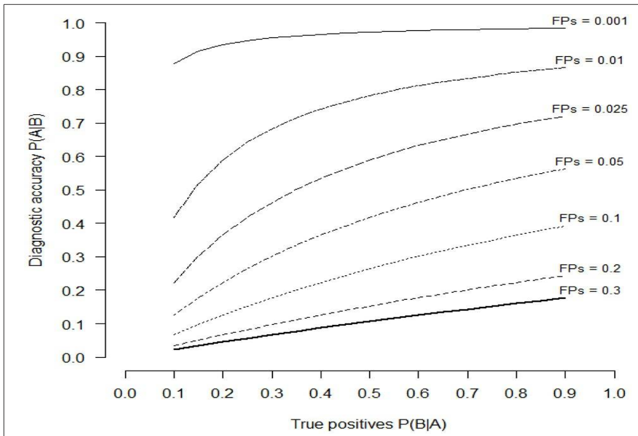


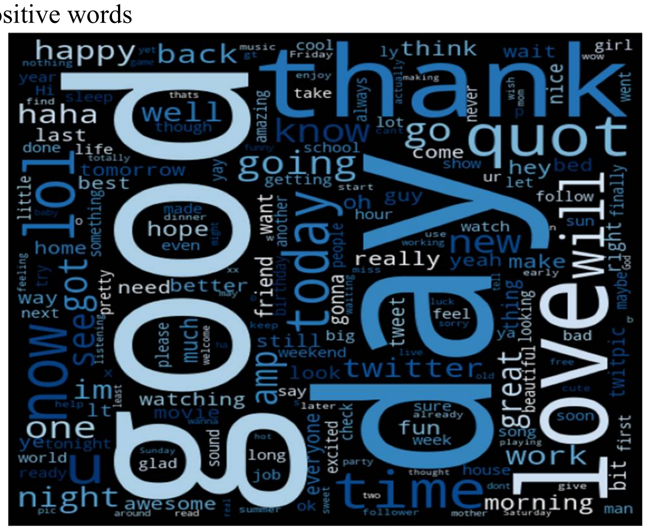
FIGURE 1 | Diagnostic accuracy of an hypothetical psychological test in relation to the different levels of TPs (P(A|B) values of the horizontal axis) and of FPs (series of curves with different line styles). Diagnostic accuracy is estimated with the Bayes' theorem (Equation 1). The selected levels of FPs were 30, 20, 10, 5, 2.5, 1, and 0.1%.

IV. WORD CLOUD

Word Cloud is a technique of visualization of data which represents text data according to the size of each word indicating the frequency word cloud is mainly used to examine the particular highlighted words or data from social media sites.

The modules mainly used in word cloud are Matplotlib, word cloud, and pandas . Word cloud holds some drawback too and that is the accuracy of translating a font size is impossible, Word cloud are not perfect in every situations. Apart from negative drawbacks word cloud holds some positive feedback too that is it is easy to understand for new comers, it takes and analyse the negative and positive words and predict the highest accuracy of correct patterns, it identifies and translates the new key phrases and SEO.

1) Case 1 . Word cloud data on positive words

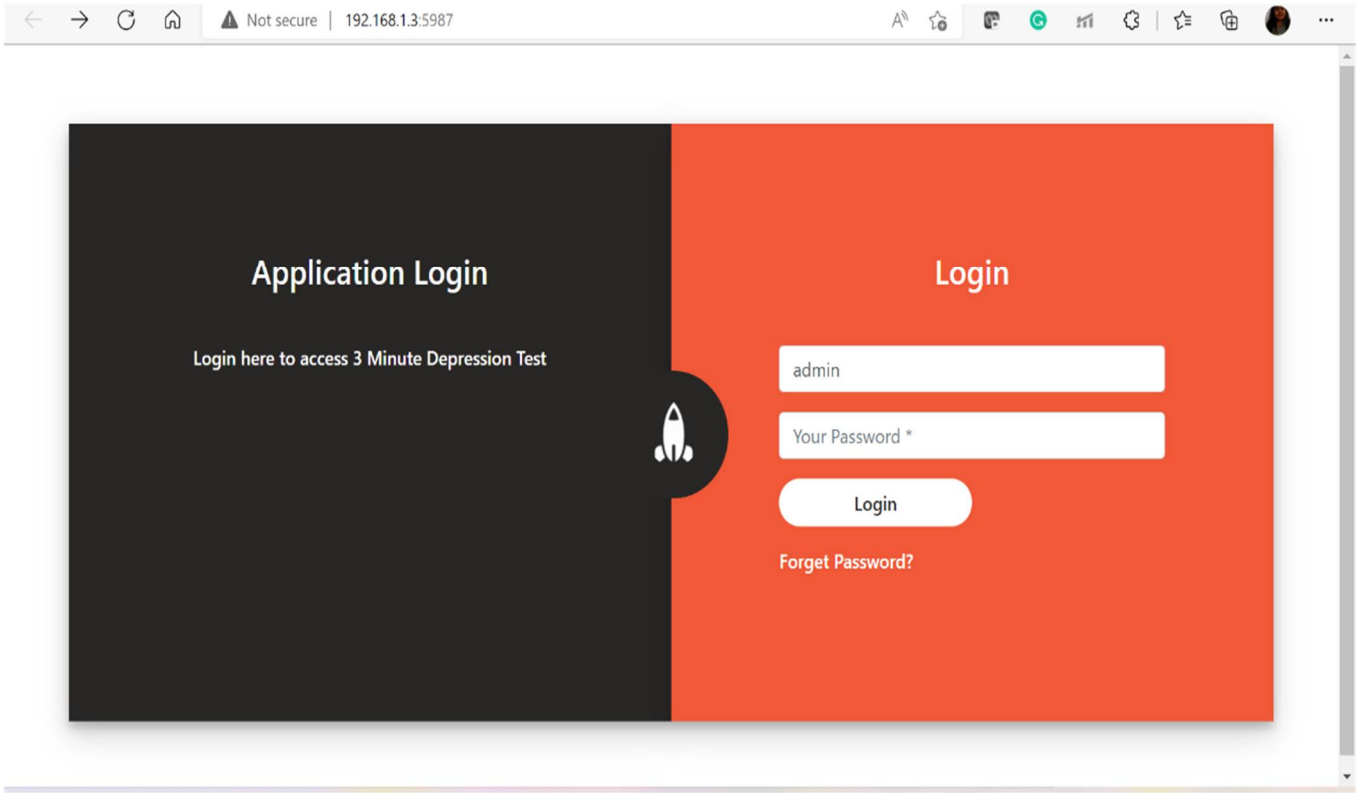


2) Case 2 . Word cloud data on negative words

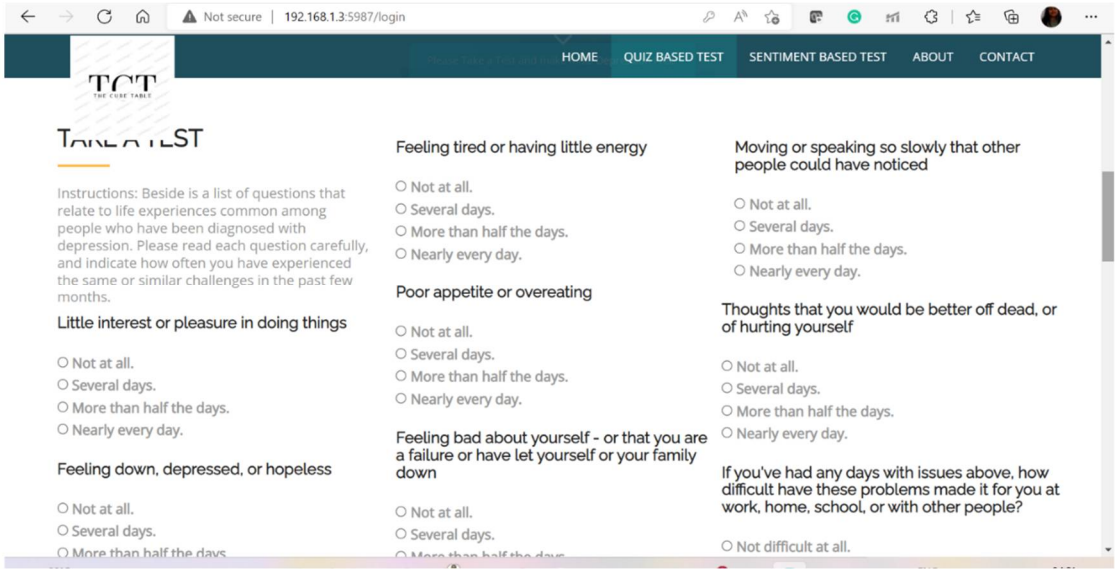


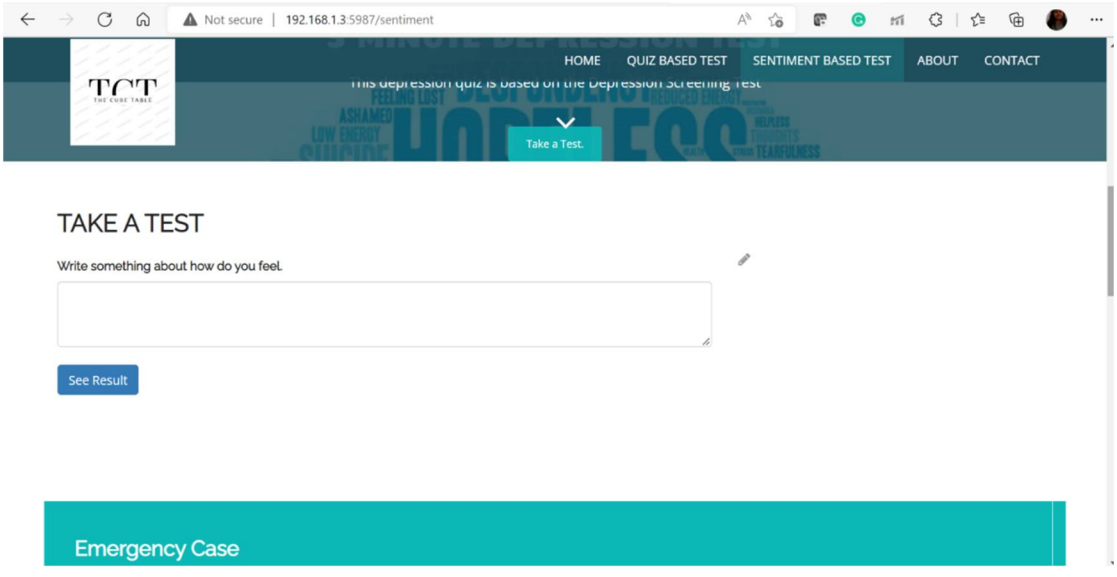
V. RESULT AND ANALYSIS

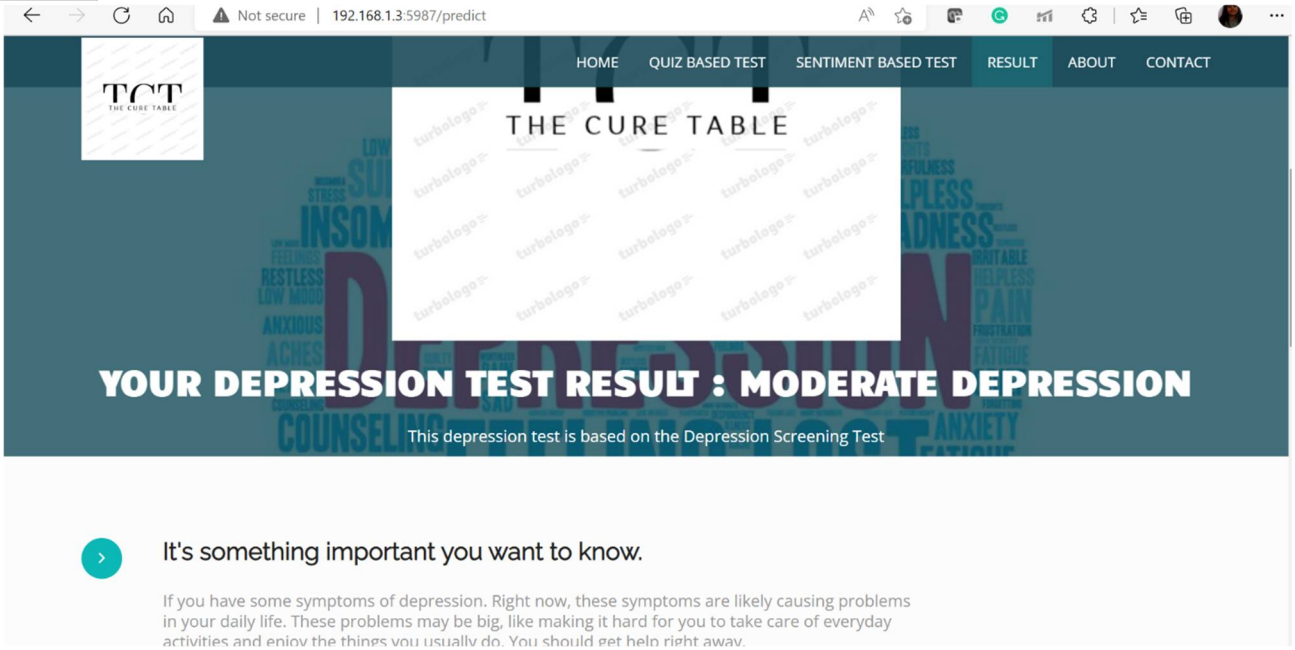
In this paper, we have reviewed the techniques that can be applied for depression and anxiety based analysis. We applied decision tree, KNN based, SVM, for depression depression of emotional terms. We have a built a user interface that can be utilized to identify the early signs




HOME
QUIZ BASED TEST
SENTIMENT BASED TEST
ABOUT
CONTACT
TCT
THE CURE TABLE


Not secure | 192.168.1.3:5987/login
HOME
QUIZ BASED TEST
SENTIMENT BASED TEST
ABOUT
CONTACT
Instructions: Beside is a list of questions that relate to life experiences common among people who have been diagnosed with depression. Please read each question carefully, and indicate how often you have experienced the same or similar challenges in the past few months.
Little interest or pleasure in doing things
Not at all.
Several days.
More than half the days.
Nearly every day.
Feeling down, depressed, or hopeless
Not at all.
Several days.
Feeling tired or having little energy
Not at all.
Several days.
More than half the days.
Nearly every day.
Poor appetite or overeating
Not at all.
Several days.
More than half the days.
Nearly every day.
Feeling bad about yourself - or that you are a failure or have let yourself or your family down
Not at all.
Several days.
Moving or speaking so slowly that other people could have noticed
Not at all.
Several days.
More than half the days.
Nearly every day.
Thoughts that you would be better off dead, or of hurting yourself
Not at all.
Several days.
More than half the days.
Nearly every day.
If you've had any days with issues above, how difficult have these problems made it for you at work, home, school, or with other people?
Not difficult at all.


Not secure | 192.168.1.3:5987/sentiment
HOME
QUIZ BASED TEST
SENTIMENT BASED TEST
ABOUT
CONTACT
Take a Test.
TAKE A TEST
Write something about how do you feel.
See Result
Emergency Case



VI. PERFORMANCE ANALYSIS

Though the result is generated from only a single algorithm, we test four different algorithms in the model to compare their accuracies. These classification algorithms are Naïve Bayes Algorithm, Decision Tree Algorithm, KNN Algorithm and Random Forest Algorithm. In the proposed system, the Decision Tree Algorithm is the most efficient algorithm with an accuracy of 97%, with Random Forest Algorithm being the least efficient with an accuracy of 40%.

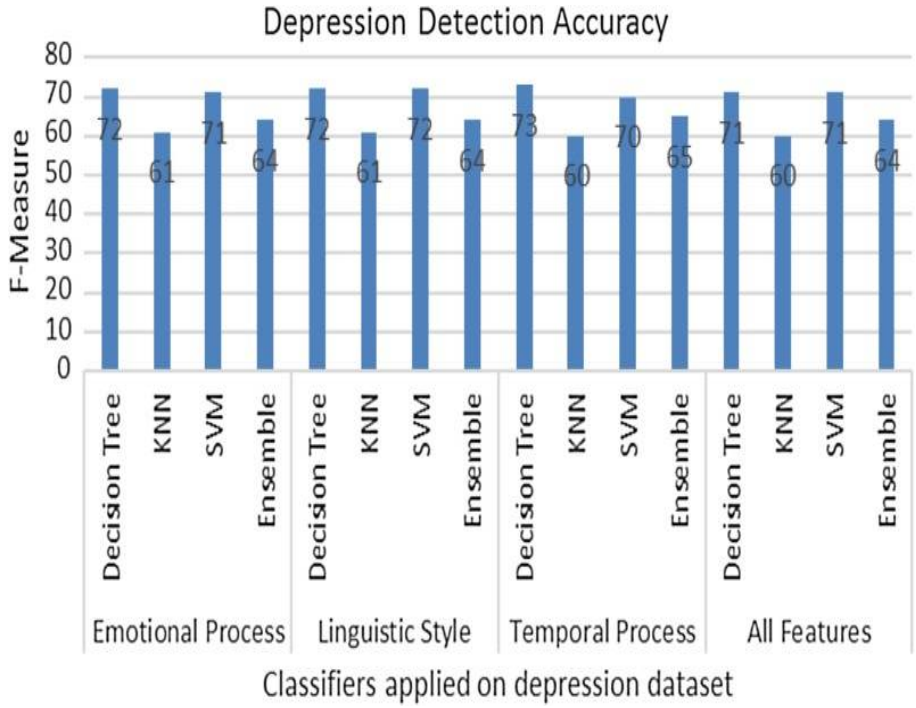


Figure 3

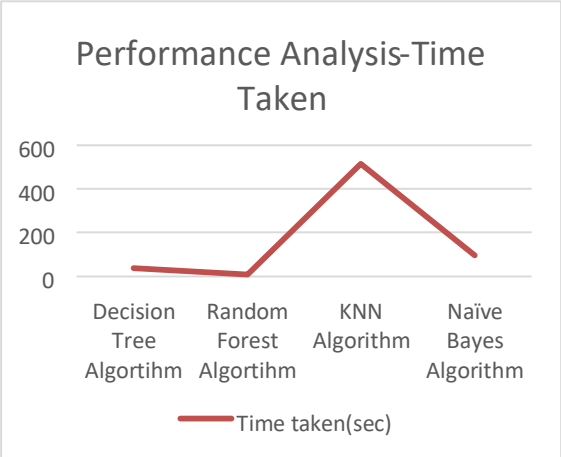


Figure 4

VII. CONCLUSION AND FUTURE WORK

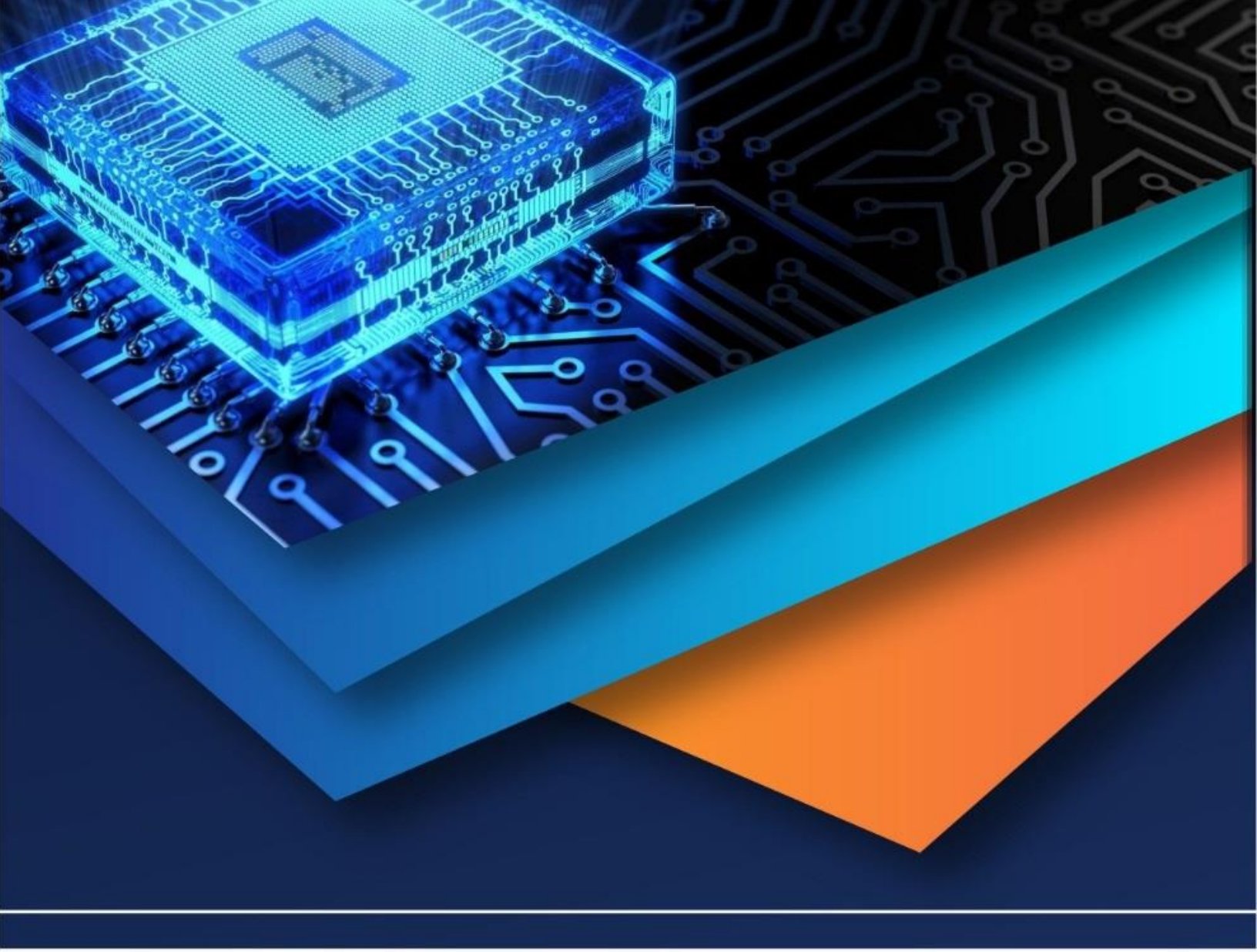
Proposed system aims upon reducing physical intervention of the human beings in the process of detecting depression and anxiety based signs of early detection on particular individual .The provided diagnostic comes with high accuracy
level, owing to the larger dataset on which model has been trained upon.

The future development on the proposed system shall be focused on making the application more reachable to the people, will make the model more accurate and precise. This could be achieved by adding different languages for diagnosis, mainly the native languages spoken in India.

Further improvements shall be made by moving out of the text classification model, and adapting to voice recognition, facial expressions model that can detect depression in various languages with varied multiple ways.

REFERENCES

[1] Aaron T Beck, Robert A Steer, and Gregory K Brown, (1996), "Beck depression inventory-ii", San Antonio,vol. 78no. 2,pp. 490–8.
[2] A. Benton, G. Coppersmith, and M. Dredze, "Ethical research protocols for social media health research," in Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, EthNLP@EACL, Valencia, Spain, pp. 94–102,2017.
[3] Akkapon Wongkoblap, Miguel A. Vadillo and Vasa Curcin, "Classifying Depressed Users with Multiple Instance Learning from Social Network Data", IEEE International Conference on Healthcare Informatics (ICHI), Vol.1 Issue.5, pg. 611-621, 2018.
[4] Alex J Mitchell, Sanjay Rao, and Amol Vaze, "International comparison of clinicians' ability to identify depression in primary care: metaanalysis and meta-regression of predictors". Br J Gen Pract, 61(583): e72–e80, 2011.
[5] Andrew Yates, Arman Cohan, and Nazli Goharian, " Depression and self-harm risk assessment in online Forums", In The Conference on Empirical Methods in Natural Language Processing, pp. 2968–2979, 2017.
[6] Arkaprabha Sau and Ishita Bhakta, "Predicting anxiety and depression in elderly patients using machine learning technology", Healthcare Technology Letters, Vol.4 Issue.6, pp. 238 – 243, 2017.
[7] B. Pang and L. Lee, "Opinion mining and sentiment analysis,"Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp.1–135.
[8] Carol Roeloffs, Cathy Sherbourne, J¨urgen Un¨utzer, Arlene Fink, Lingqi Tang, and Kenneth BWells (2003), "Stigma and depression among primary care patients", General hospital psychiatry, 25(5):311–315, 2018.
[9] Daniel Eisenberg, Marilyn F Downs, Ezra Golberstein, and Kara Zivin, "Stigma and help seeking for mental health among college students", Medical Care Research and Review, vol. 66, no. 5, pp. 522–541, 2009.
[10] Grohol, J. M, Using websites, blogs and wikis in mental health. In K. Anthony, D. A. N. Nagel, and S. Goss (eds.), "The use of technology in mental health applications ethics and practice", Springfield, IL: Charles C. Thomas, pp. 68–75, 2010.
[11] JT Wolohan, Misato Hiraga, Atreyee Mukherjee and Zeeshan Ali Sayyed, "Detecting Linguistic Traces of Depression in Topic-Restricted Text: Attending to Self-Stigmatized Depression with NLP", First International Workshop on Language Cognition and Computational Models, Vol.3 Issue.5, pp. 11-21, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)