iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51067

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 11 Issue V May 2023- Available at www.ijraset.com

Depression Detection and Analysis using ML

Prof. Dr. Rohini Temkar[1], Suhail Shaikh[2], Shalini Mirani[3], Sakshi Patil[4], Siyona Singh[5]
Department of Computer Engineering, VES Institute of Technology(University of Mumbai) Mumbai, India

Abstract: Physiological interference in a person's life can bring out many difficulties which affect the basic abilities of a person to do simple tasks. Depression, one such issue, can be detected through only medically trained psychiatrists and the early detection of depression is crucial in providing effective treatment. Social Media platforms prove to be a valuable source of information for this motive. Social Media provides a safe platform for individuals to express their emotions and feelings while at the same time, it can also contribute to the development of depression symptoms particularly in vulnerable populations i.e. Younger generations, due to factors such as social comparison, cyberbullying, social isolation and constant need for validation. This paper further explores evidence of the link between social media use and depression in aiding early detection. The objective of the current study is to apply various ML techniques to assist psychiatrists in recognizing patient symptoms. The plat-form used for our research is Twitter. The model detects the symptoms/behavior of early depression in the users through their current and past tweets. For this purpose, we have deployed various approaches to train and test an ML model or classifier using the right features depending on the information gathered from a questionnaire and through the features extracted from a user's tweets and their social network activities.

Keywords: Depression, Social Media, Psychiatrists

I. INTRODUCTION

Depression Analysis using ML is a machine learning model which is trained to predict various depressive disorders. The user has to answer some of the questions which are framed particularly regarding the matter and the system will use a trained data set and different models to predict depression. If the detection of depression is still complex for trained practitioners or psychiatrists, further implementation of detect-ing depression through social media platforms like Twitter is applied. The project's goal is to assist in the worthy cause of identifying and treating mental health illnesses, including depression-related disorders of many subtypes that affect people of all ages, from children to seniors. We analyze various extracted features through effective machine-learning algorithms to make the final statement. In the problem state-ment, we have used publicly available tweets containing the patient's tweets to classify them accordingly. In the study, we analyze various cues to detect the emotional events: the place of cause event and experience relative to the emotion keyword i.e. positive emotions like ('happy', 'good', 'nice', etc), negative emotions('worthless','ugly', 'useless', etc) and various other keywords were sorted as per the emotions exhibited accordingly.

A. List of Abbreviations

1) ML - Machine Learning
2) SVC - Support Vector Classification
3) LR - Logisitic Regression

II. LITERATURE SURVEY

1) Authors: Md.Sabab Zulfiker, Nasrin Kabir, Al Amin Biswas, Tahmina Nazeen, and Mohammad Shorif Uddi
Abstract: This model have predicted depression by find-ing the common factors of depression using 604 partic-ipants. They obtained an accuracy of 92.56

2) Authors: Sonam Gupta, Lipika Goel, Arjun Singh, Ajay Prasad, and Mohammad Aman Ulla
Abstract: This paper uses social media platforms to predict depression by collecting customers' opin-ions(positive, negative, and neutral) for a product or any activity. The limitation of their study is that their model will not be able to help many people who do not use social media leaving the people undiagnosed.

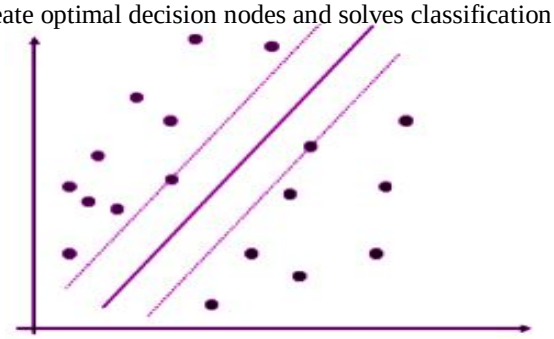
3) Authors: Md.Rafiqul Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M. Kamal, Hua Wang and Anwarr Ulhaq
Abstract: in [3] studied various signs of depression on Facebook and used them to predict depression among Facebook users. This was done by studying the emo-tional process, temporal process, and linguistic style factors and training a model to utilize each type of factor. Their model had the highest accuracy when they used the Decision Tree(DT) ML approach.

4) Authors: Umme Marzia Haque, Enamul Kabir, Rasheda Khanam
Abstract: This paper aim at detecting depression accu-rately using Random Forest (RF) in children aged 4-17 years. They have extracted features accurately using correlation and weighted classifiers item
Authors: Jini Jojo Stephen,Prabu P
Abstract: This model aims at detecting the level of depression in Twitter users. Their future goal is to upgrade their model by checking the patient's activity on all social media platforms to detect her/his level of depression more accurately. item
Authors: Nisha Shetty, Balachandra Muniyal, Arshia Anand, Sushant Kumar, Sushant Prabhu
Abstract: They have used sentiment analysis to predict depression in Twitter users.

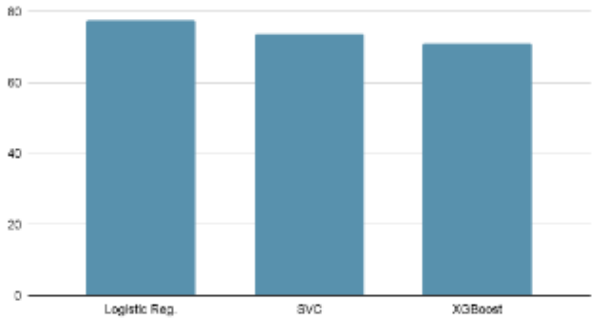
a) In the existing system no frontend and ui is present if present very limited. Because of this, normal people or non-medical people are not able to use those systems .
b) The accuracy of the existing system is low, not greater than 90%. This low accuracy makes decisions uncertain for users. In today's world we need higher accuracy because heart disease is a very complex disease.
c) The data set is low (303 data set). Because of less data it will be difficult for the system to predict whether a person has heart disease or not.
d) In the present time new and powerful machine algorithms are present. But existing systems use a very limited num-ber of machine learning algorithms. These are one of the reasons for low accuracy of existing systems. Algorithms are LMT, Random forest, Decision tree, KNN, Naive Bayes, Logistic Regression, SVM.

III. METHODOLOGY

1) *Data Collection:* For the effective purpose of detecting depression, the dataset is collected from the question-naire provided to the patients and through websites like Twitter. First, the dataset is collected and cleaned. Secondly, through the cleaning process, we are pro-vided with some important keywords/features. If the tweets or the answers from the questionnaire do not contain the features extracted from the dataset they are grouped as non-depressive. Further to distinguish between depressive and nondepressive tweets some im-portant features were selected. Accordingly, to maxi-mize the modeling performance of our model various Machine Learning Algorithms are used like XGB Classifier, Random Forest Classifier, Logistic Regression, Support Vector Machine(SVM) and Random Forest Classifica-tion. Currently, Twitter API is to be used in order to extract real users tweets that are active to this day.
2) *Data Filtering:* Filtering is a preprocessing step to fil-ter out any redundancies that the input dataset con-tains. The dataset provided by the questionnaire and the tweets are filtered out to generate stop words such as ['i',' me','my',' myself', 'we',' our', 'ours',' ourselves',' you,"you're","you've"," you'll"," you'd",' your',' yours',' yourself', etc] which do not provide any meaning to our detection model. The final aim of this process is to generate a large cleaned dataset without any redundancies and missing values .
3) *XGB Classifier:* XGBoost (Extreme Gradient Boosting) is able to handle real - world dataset with missing values aiming to build a strong classifier on the basis of the number of weak classifiers. Gradient Boosted trees, in which each predictor corrects the inaccuracy of its predecessor, are the basis for XGB.
4) *Random Forest Classifier:* Random Forest is a flexible al-gorithm which tackles both classification and regression in the model, reaching the goal node based on multiple states of a decision tree.
5) *Logistic Regression:* Logistic regression is a statistical technique used when we have to describe the relation-ship between a dependent variable and one or more independent variables and data, predicting the finite number of the outcomes.
6) *Support Vector Machine (SVM):* It is a supervised ma-chine learning algorithm, it creates a decision boundary or chooses extreme points in the dataset to create optimal decision nodes and solves classification problems.



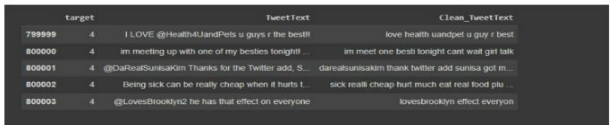
7) *Measuring Accuracy of Model:* Measuring the accuracy of a model is an essential step in evaluating its performance. A common method for measuring accuracy is cross-validation. It is a statistical technique that involves dividing the data into multiple subsets, and training the model on each subset while using the other subsets for testing. The data is divided into k equally sized subsets. Then the model is trained on k-1 subsets and tested on the remaining ones, and the process is repeated k times so that each subset is used for testing once. The accuracy of the model is calculated by averaging the results of each iteration. The method also proves to be a more robust estimate accuracy model as compared to a single train-test split, as it reduces the risk of over fitting or under fitting due to the random selection of training and testing data. To summarise, cross-validation is one of the methods used to measure the accuracy of the model by dividing the data into multiple subsets and training and testing the model on each subset.

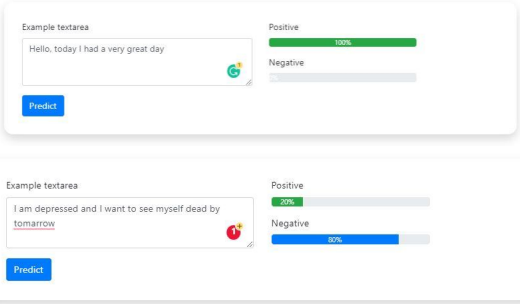


A. Data and Results

The model verifies the features provided by the patient and if necessary the doctor can make use of the patients social media activity and their tweets which the model shall take as input then after going through cleaning ,processing and removal of URLs(if any) the necessary features are extracted and the model takes half of them for training and the rest for testing the data set.

Fig





Metrics and Evaluation: Our main goal is to predict the accuracy for future problems that the disease may cause and which algorithm gives more accuracy that can be made for the target output counts that a person has Heart Disease or not. Because our project is a classification problem, we evaluate the models using accuracy, precision, recall, and F1 scores.

IV. ACKNOWLEDGMENT

We are thankful to our college Vivekanand Education So-ciety's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project. It gives us immense pleasure to express our deep and sincere gratitude to Professor Mrs. Rohini Temkar (Project Guide) for her kind help and valuable advice during the development of the project and for her guidance. We are deeply indebted to the Head of the Computer Department Dr.Nupur Giri and our Principal Dr.J.M. Nair for giving us this valuable opportunity to do this pride. We sincerely thank them for their cooperation and their assistance without which we would have struggled to complete this project overview and project review satisfactorily. We would like to express our heartfelt appreciation to all teaching and non-teaching personnel for their consistent encourage-ment, support, and unselfish assistance throughout the project work. It gives me great pleasure to recognize the Department of Computer Engineering's assistance and suggestions. We would like to offer our heartfelt gratitude to everyone who assisted us in acquiring project information. Our families, too, have supplied moral support and encouragement on numerous occasions

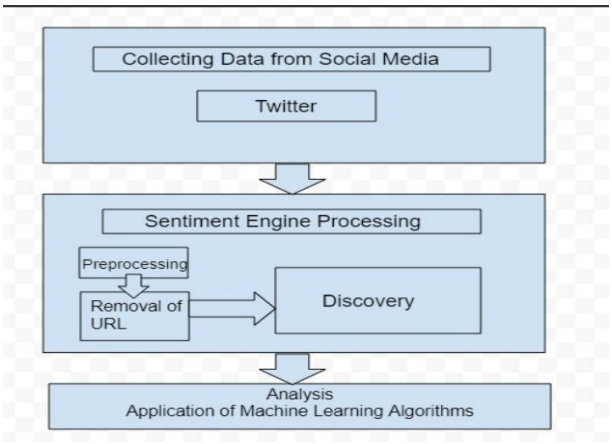


Fig. 1. Flowchart

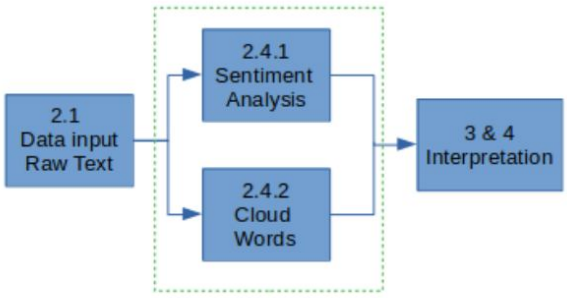
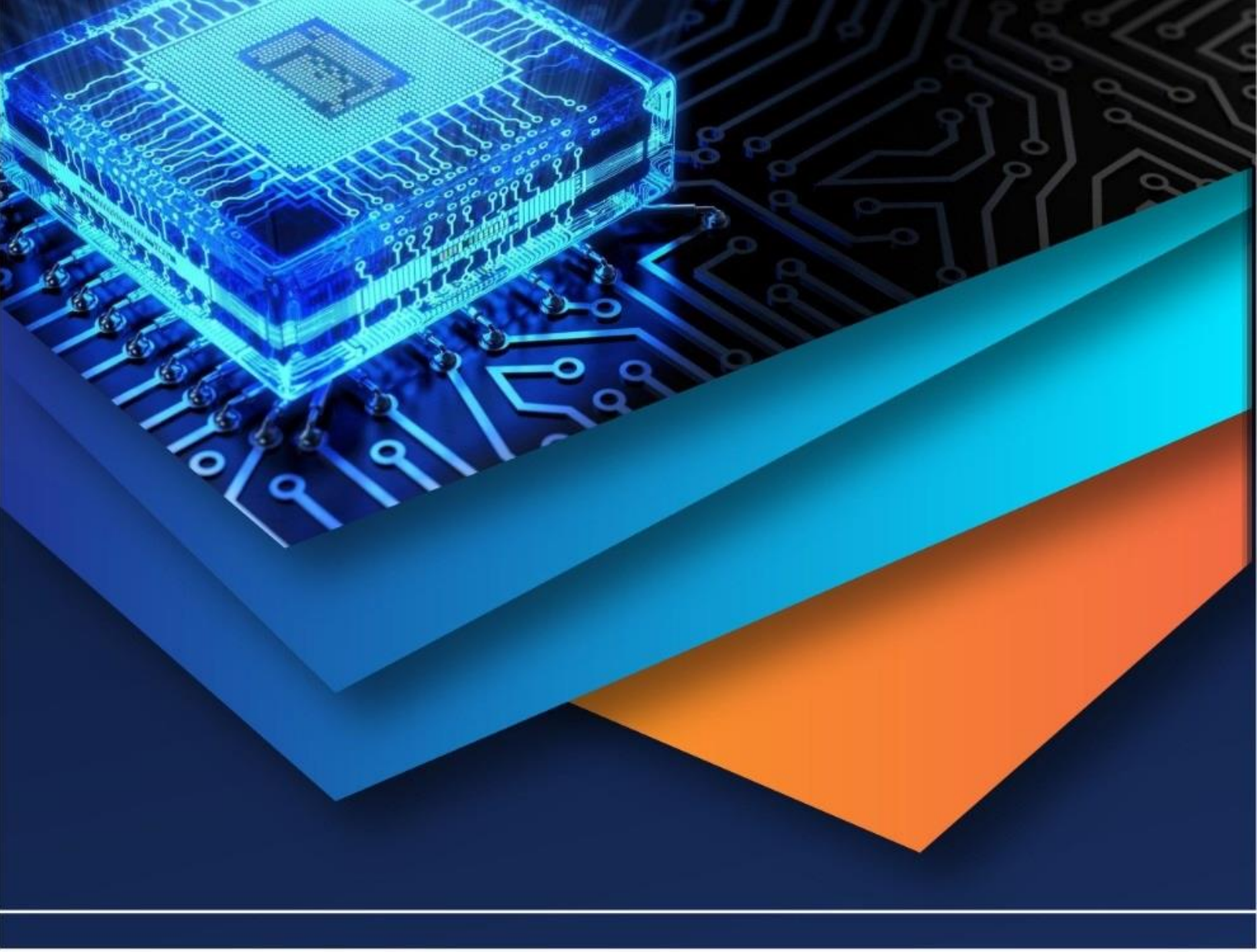


Fig. 2. Block Diagram

REFERENCES

[1] Md.Sabab Zulfiker, Nasrin Kabir, Al Amin Biswas, Tahmina Nazeen, Mohammad Shorif Uddin(2021) An in-depth analysis of machine learning approaches to predict depression.
[2] Sonam Gupta, Lipika Goel, Arjun Singh, Ajay Prasad, and Mohammad Aman Ullah(2022) Psychological Analysis for Depression Detection from Social Network Sites
[3] Md.Rafiqul Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M. Kamal, Hua Wang and Anwarr Ulhaq(2018) Depression detection from social network data using machine learning techniques. [
[4] Umme Marzia Haque, Enamul Kabir, Rasheda Khanam(2021) Detection of child depression using machine learning methods.
[5] Ramin Safa, Peyman Bayat, Leila Moghtader(2021) Automatic detection of depression symptoms in twitter using multimodal analysis.
[6] Jini Jojo Stephen,Prabu P.(2019) Detecting the magnitude of depression in Twitter users using sentiment analysis
[7] Nisha Shetty, Balachandra Muniyal, Arshia Anand, Sushant Kumar, Sushant Prabhu(2020) Predicting depression using deep learning and ensemble algorithms on raw twitter data.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY