

DEPRESSION DETECTION USING MACHINE LEARNING

KULKARNI MRUNMAYEE¹, KHARADE PRANOTI², BHALERAO ASAVARI³, Prof. CHAVAN SUNITA⁴

Student, Adsul's Technical Campus Faculty of Engineering, Chas, Ahmednagar^{1,2,3}

Asst. Prof. Adsul's Technical Campus Faculty of Engineering, Chas, Ahmednagar⁴

Abstract—Currently, over 300 million people worldwide suffer from depression disorders to varying degrees. Including social stigma, treatment cost and availability, might prevent affected individuals from seeking help. 60 percent of those affected by mental illnesses do not receive treatment. So to overcome this problem the project aimed to be identifying the person suffers from depression and the root problem. As the system gets trained by the inputs given by the user, it starts customizing the recommendation and responses that address the root cause.

Keywords— *Machine Learning, depression prediction, speech to text, text to speech, Data cleaning, Data Normalization, Data collection, SVM classifier*

I. INTRODUCTION

Major depressive disorders is a disorder affecting person's way of living. Depression leads to negative affect and mood. It has such an impact on people that they started feeling ore of those leisure activities which they used to enjoy earlier. Their approach is negative towards everything. This is a common disorder for every generation. But millennial report higher rates of depression than any other generation. About 20% of the people suffer from depression and anxiety. They feel that stress keeps them awake all night (Health Status). Machine Learning is getting computers to program themselves. It is automating the process of automation. Machine Learning is used in computer science and other fields. And so, for this type of intelligent automation, machine should be learned with complex patterns as that of human brain. There are various methods used for prediction of depression such as facial images, speech tones etc. But most of the ongoing researches are based on a particular learning method. To know which prediction model is accurate a comparison between different models is necessary. This research work focuses on study of three major methods used for prediction of depression: - Using Machine Learning Classifiers and WEKA -Using Imaging and Machine Learning Methods -Using Risk Factors And to detect the most consistent and accurate of them all. Also a software developed by University of Waikato, New Zealand called WEKA, is used in one of the above prediction model. It is used for classification and depression among people. Rest of the sections are classified as: Section II gives the overview of the different prediction algorithm for depression. Section III provides the observation from the past studies Section 4 provides a

basic discussion of the outcomes derived from the past studies. Section 5 concludes the research work.

Motivation

This system will help to detect the depression among the people. This system helps to measure the percentage of the depression using voice and text. Motivation of this project is to detect the depression at early stage before it gets affects human life and reduce depression level using multi classifier system.

Scope

Scope of this project is to detect depression which is based on different category and life issue and physical health. New machine . New machine learning method could be deployed in smartphones. Making mental health care cheaper and more accessible. The system will work efficiently for the short interviews/test to detect the basic mental condition. Though it is a software device so there is no need to the person to come over the place, the interview/test can be conducted on social platform. learning method could be deployed in smartphones.

II. LITERATURE SURVEY

Major depressive disorder, is the most frequent mood disorder, highly concurrent with anxiety and stress. It has an estimated 17% of lifetime prevalence. It causes major effects on quality of life, healthcare utilization etc. MRI (Magnetic Resonance Imaging) uses high frequency radio waves to produce images of a particular body organ for its detailed examination. MRI modalities to a major extent are used in depression.

| Sr. No. | Title | Author | Using Machine learning methods | Accuracy |
|---------|---|-------------------------|---|---------------------|
| 1. | Prognostic and diagnostic potential of the structural neuro anatomy of depression | Costafreda et al., 2009 | Support vector machines | 67% |
| 2. | Multi-center diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder | Mwangi et al., 2012 | Support vector machines (nonlinear Gaussian kernel) - Relevance vector machines (nonlinear Gaussian kernel) | Svm 87.1% Rvm 90.3% |
| 3. | Resting-state functional connectivity bias of middle temporal gyrus and caudate with altered gray matter volume in major depression | Ma, Chaoqiong, et al. | Support vector machines (linear kernel) | 94.3% |

These are used to study brain structure from different aspects. Most common of these modalities are: T1-weighted imaging is used to study critical regions T2-weighted imaging is used to study white matter hyper intensities DTI (Diffusion Tensor Imaging) is used to get understanding of the brain from a microscopic level. To study the diffusion of molecules in the brain tissues.

Machine learning here is used in a unique way to test the potential of each MRI measure as a relevant biomarker of depression. Depending upon the data there are 3 types of learning: o SL – If all data is labelled o SSL - If there is labelled data along with unlabeled o USL - If there is unlabeled data only In supervised learning, training data includes desired outputs. It works only for labelled data. In SSL, some desired outputs are considered in the training data. It works for labelled data along with unlabeled data. In unsupervised learning, training data doesn't include desired output. It works for unlabeled data. The SL and SSL are divided into 2 methods – a. Classification Based Method – Under this method it classifies the data according to discrete categorical labels.

III. RELATED THEORY AND PROBLEM DEFINITION

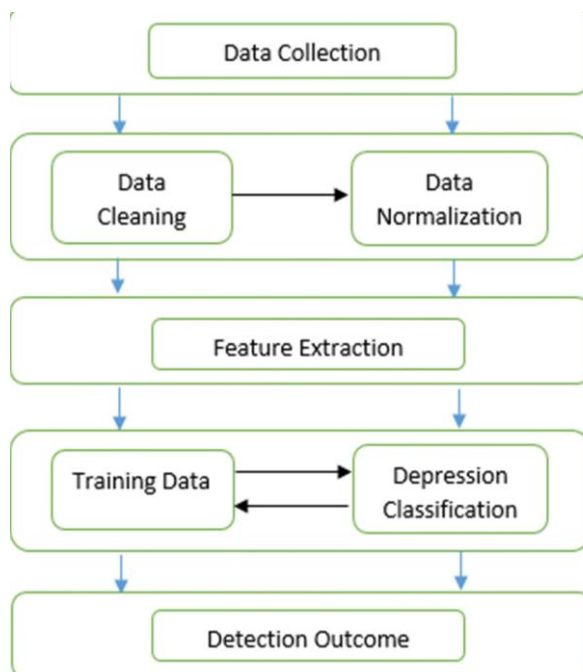
Problem Definition

Changing life cycle and the increasing competition in each field causes the fear in human's life to interact with each other. Also, it affects in human nature that the person won't even be able to express his feelings which affects in the day today's life of the person. Depression is technically a mental disorder, but it also affects your physical health and well-being. Learn more about some of the most common symptoms of depression, as well as how depression can affect your entire body, especially if left untreated. Feeling sad or anxious at times is a normal part of life. but if these feelings last more than two weeks they could be symptoms of depression. It's estimated that each year 17 million American adults will experience depression. However, clinical depression, especially left untreated, can interrupt your day-to-day life and cause a ripple effect of additional symptoms. Depression affects how you feel and can also

cause changes in your body. Major depression (a more advanced form of depression) is considered a serious medical condition that may have a dramatic effect on your quality of life. Depression may be more difficult to detect in children who can't articulate their symptoms. Behaviors you may want to look out for include persistent clinginess, worry, and unwillingness to attend school without improvement over time. Children may also be excessively irritable and negative.

System Architecture

In this section we discuss the application layout, architecture and the components they are composed of. We also describe the communication that takes place within the different modules and identify the website feature classification. Along the way we intend to present a general model of an application which we will refer throughout the thesis. This enables us to study our problem from general perspective, without having to consider architecture or implementation specification details.



Data collection:

Data collection is the process of gathering and measuring information from countless different sources. In order to use the data, we collect develop practical artificial intelligence and machine learning solutions.

Data Cleaning:

Today data scientists often end up spending 60% of their time cleaning dirty data before they can apply analytics or ML. Data cleaning is essentially the task of removing errors and anomalies or replacing observed value with true values from data to get more value in analytics.

Data Normalization:

Normalization is a technique often applies as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to a common scale ,without distorting differences in the ranges of values.

Feature Extraction:

In machine learning, pattern recognition and in image processing ,feature extraction starts from an initial measured data and builds derived values(features)intended to be informative and non-redundant ,facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations

Training Data:

The training data is an initial set of data used to help a program understand how to apply technologies like neural networks to learn and produce sophisticated results. Taring data also known as training set, training dataset or learning set.

3.1 Analysis Models: SDLC Model to be applied

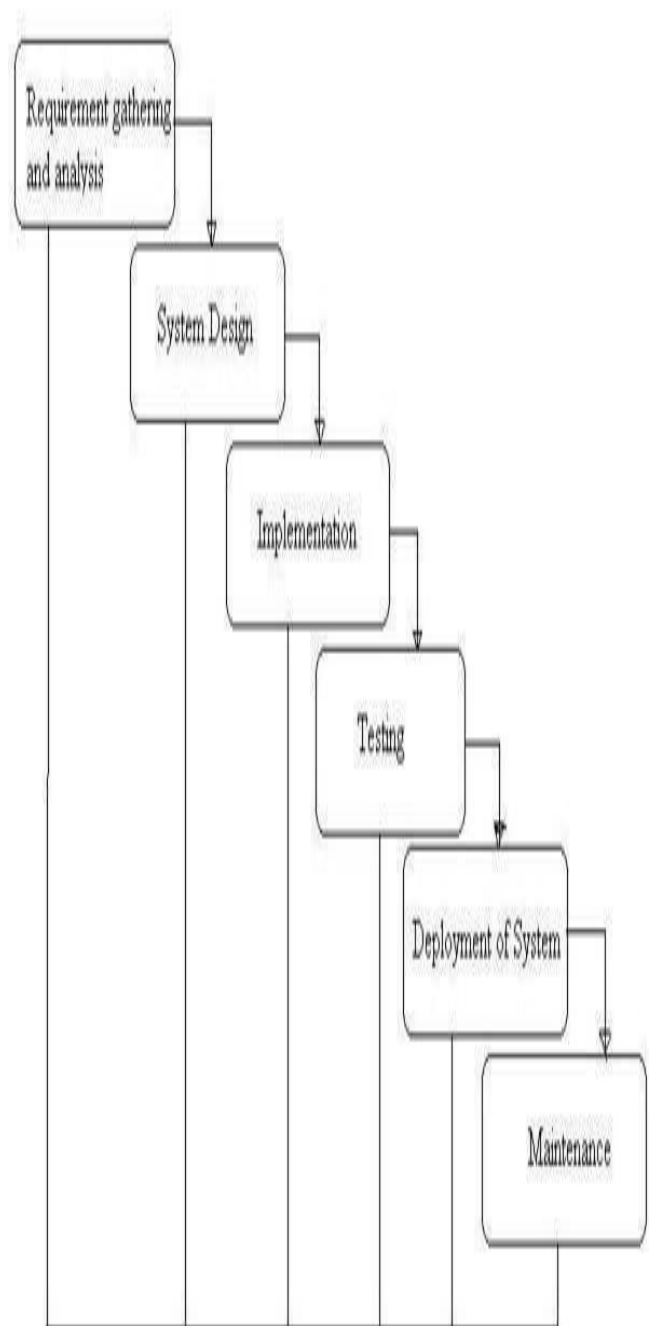
The Waterfall Model is sequential design process, often used in Software development processes; where progress is seen as flowing steadily download through the phase of conception, Initiation, Analysis, Design, Construction, Testing, Production/Implementation and Maintenance. This Model is also called as the classic Life cycle model as it suggests a systematic sequential approach to software developments. This one of the oldest model followed in software engineering. The process begins with the communication phase where the customer specifies the requirements and then progress through other phases like planning, modeling, construction and deployment of the software.

There are 5 Phase of waterfall model:

The Waterfall Model is sequential design process, often used in Software development processes; where progress is seen as flowing steadily download through the phase of conception, Initiation, Analysis, Design, Construction, Testing, Production/Implementation and Maintenance. This Models I so called as the classic Life cycle model as it suggests a systematic sequential approach to software

developments. This one of the oldest model followed in software engineering. The process begins with the communication phase where the customer specifies the requirements and then progress through other phases like planning, modeling construction and deployment of the software.

General Overview of "Waterfall Model"



COMMUNICATION

In communication phase the major task performed is requirement gathering which helps in finding out exact need of customer. Once all the needs of the customer are gathered the next step is planning.

PLANNING

In planning major activities like planning for schedule, keeping tracks on the processes and the estimation related to the project are done. Planning is even used to find the types of risks involved throughout the projects. Planning describe show technical tasks are going to take place and what resources are needed and how to use them.

MODELING

This is one the important phases as the architecture of the system is designed in this phase .Analysis is carried out and depending on the analysis a software model is designed. Different models for develop software are created depending on the requirements gathered in the first phase and the planning done in the second phase.

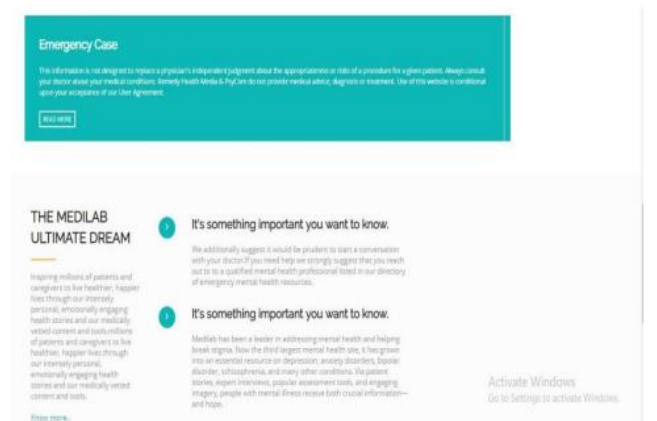
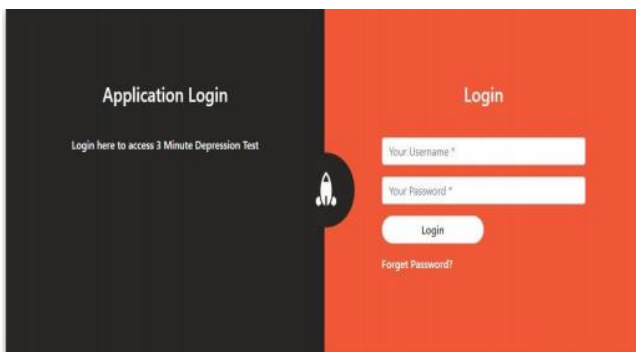
CONSTRUCTION

The actual coding of the software is done in this phase. This coding is done on the basis of the model designed in the modeling phase. So in this phase software is actually developed and tested.

DEPLOYMENT

In this last phase the product is actually rolled out or delivered installed at customer' send and support is given if required. A feedback is taken from the customer to ensure the quality of the product. From the last two decades Waterfall model has come under lot of criticism due to its efficiency issues. So let's discuss the advantages and disadvantages of waterfall model.

IV. RESULT



CONCLUSION

Depressive disorder makes people unfit not only physically but mentally also. The past studies the three major methods were studied to determine the most accurate method of them all. After studying all these method we conclude that the most accurate and the consistent method is the Bayes net Classifier for the percentage split testing. The term "depression" is used in a number of different ways. It is often used to mean this syndrome but may refer to other mood disorders or simply to a low mood. People's conceptualizations of depression vary widely, both within and among cultures. "Because of the lack of scientific certainty," one commentator has observed, "the debate over depression turns on questions of language. What we call it—'disease,' 'disorder,' 'state of mind'—affects how we view, diagnose, and treat it. There are cultural differences in the extent to which serious depression is considered an illness requiring personal professional treatment, or is an indicator of something else, such as the need to address social or moral problems, the result of biological imbalances, or a reflection of individual differences in the understanding of distress that may reinforce feelings of powerlessness, and emotional struggle Physical exercise is recommended for management of mild depression, and has a moderate effect on symptoms. Exercise has also been found to be effective for (unipolar) major depression. It is equivalent to the use of medications or psychological therapies in most people. In older people it does appear to decrease depression. Exercise may be recommended to people who are willing, motivated, and physically healthy enough to participate in an exercise program as treatment. There is a small amount of evidence that skipping a night's sleep may improve depressive symptoms, with the effects usually showing up within a day. This effect is usually temporary. Besides sleepiness, this method can cause a side effect of mania or hypomania. In observational studies, smoking cessation has benefits in depression as large as or larger than those of medications. Besides exercise, sleep and diet may play a role in depression, and interventions in these areas may be an effective add-on to conventional methods.

REFERENCES

- [1] Blanco, Joel A., and Lynn A. Barnett. "The effects of depression on leisure: varying relationships between enjoyment, sociability, participation, and desired outcomes in college students." *Leisure Sciences* 36.5 (2014): 458-478.
- [2] Domingos, Pedro. "A few useful things to know about machine learning." *Communications of the ACM* 55.10 (2012): 78-87.
- [3] Remco R. Bouckaert, "Bayesian network classifiers in WEKA", JULY 2007.
- [4] Komarek, Paul. "Logistic regression for data mining and high dimensional classification." *Robotics Institute*

(2004): 222.

- [5] Durgesh, K. SRIVASTAVA, and B. Lekha. "Data classification using support vector machine." *Journal of Theoretical and Applied Information Technology* 12.1 (2010): 1-7.
- [6] Zimmermann, Petra, et al. "Heterogeneity of DSM-IV major depressive disorder as a consequence of subthreshold bipolarity." *Archives of general psychiatry* 66.12 (2009):1341-1352.
- [7] Patel, Meenal J., Alexander Khalaf, and Howard J. Aizenstein. "Studying depression using imaging and machine learning methods." *NeuroImage: Clinical* 10 (2016): 115-123.
- [8] Mahajan, Rajiv. "Real world data: Additional source for making clinical decisions." *International Journal of Applied and Basic Medical Research* 5.2 (2015): 82.
- [9] King, Michael, et al. "Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study." *Archives of General Psychiatry* 65.12 (2008): 1368-1376.
- [10] Bhakta, Ishita, and Arkaprabha Sau. "Prediction of Depression among Senior Citizens using Machine Learning Classifiers."