International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi :https://doi.org/10.32628/IJSRCSEIT

Data Science in Mental Health

Kadambari Mirashi, Chaitrali Mirashi, Shruti Choure, Pranjal Sagar
Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra, India

Article Info
Volume 8, Issue 2
Page Number : 429-434

Publication Issue :
March-April-2022

Article History
Accepted: 10 April 2022
Published: 22 April 2022

ABSTRACT

Mental diseases are estimated to affect 700 million people worldwide. Due to the rapid increase in mental diseases in recent years, it has become critical to better comprehend the negative consequences of mental health issues. The perceived constraints of ethical norms such as autonomy protection, consent, threat, and injury make mental health research difficult. The goal of this survey was to look for studies that used big data approaches in mental illness and treatment. To begin, numerous sorts of mental disease are discussed, such as bipolar disorder, depression, and personality disorders. Mental health's impact on user behaviour, such as suicide and drug addiction, is highlighted. A discussion of the approaches and tools used to anticipate the patient's mental state using artificial intelligence and machine learning is offered.

Keywords – Support Vector Machine(SVM), Artificial Intelligence (AI), Data Analysis(DA), Pre-processing, Feature Extraction, Machine Learning(ML), Multi-Model Analysis .

I. INTRODUCTION

The project utilizes a combination of technologies and requires a set of functionalities for smooth execution and establishing accurate results. For accurate results it makes use of multiple machine learning models for classification purposes. A data science project's beginning marks with the word 'data' itself. Abundant amount of data is required for more accurate and effectual results or to be more particular 'predictions'. Moreover in context of medical data where the possible outcomes are very diverse in nature, thus more the case studies more accurate and sharp would the results be. Data collection is of prime importance, finding the right type of data is similar to finding needle in a haystack. Given the fact that in mental health sector there are itself 300+ types of subsectors each dealing with particular group of community based on age, religion, financial status and even gender and also dealing with certain type of disorder, be it MPD, Bipolar disorder, depression, phobias or even negligible anxiety disorders. Any project covering the major types of disorders poise to be itself a boon as the medication process or therapy gets a specific direction. Then comes the analysis phase wherein categorizing of symptoms, state of mind needed to be mapped with possible type of ailment. Creating and training a predictive model for accurate results demands of solid state data, but receives very distinct cases as well, where the outliers impose an obstacle but is as important to the model. Meanwhile on testing of the data although errors are to be

witnessed, the radius of possibilities definitely converges to focus on specific type of test cases, thus making it better for patients as well as practitioners. Forming an effective assessment module to generate relatable results that correspond to underlying symptoms is of prime importance. To establish the self-understanding of user and a specialized practitioner that might use the results is another factor that is considered. Furthermore, diagnosis-based procedures have the unintended consequence of preventing those who are ill from participating. As a result, psychological problems are frequently ignored or overlooked.

II. METHODS AND MATERIAL

Assumption analysis is a critical component of the project where data plays an important role. To reduce the risk of erroneous results the following pose as the main assumption and dependency. The dataset acquired and polished for prediction either gives accurate results or presents results in such a way that it gives appropriate direction to the research. Thus the predictive model is almost perfect with only a few outlier chances of wrong assumptions which would be automatically identified on going further in the process.

The system is user friendly and accessible with ease, and the results actually help in reducing the trial and error approach of the psychiatrists. The system is also user friendly and accessible to the general user and with its responsive capabilities gives appropriate results to the user. Its predictive model is susceptible to given inputs and moves forward with precision.

The system's two level approach aims to achieve perfection and accuracy, given the fact that the user or patient must the answer the questions seriously. So it is suggested to use the system under the guidance of a practitioner or experienced person. Availing to contradictory features might further give relative results but might mislead the investigation if the test is pursued for professional medication survey. It is mandatory to appear for the assessment in complete honesty as it would help the medication survey attain the correct pathway to begin medication methodologies at a therapists end.

The level2 architecture's objective is to focus on specific area of diagnosis and eliminate the non-possible type of disorder. The approach requires the data to be through multiple models to acquire better results. Applying multiple models guarantee genuineness of the system thus cross validating the results. The foundation level if cross validated, ensures the survey going in appropriate direction. It is assumed that user specifies the symptoms that are actually liable.

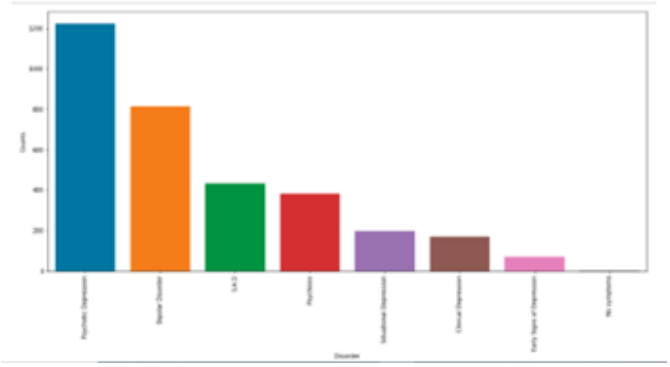


Fig.1.Graph of types of diseases

III. DATASET

Quiz1 dataset, it comprises of mixture of symptoms and binary (0, 1) flags that form patterns and sparsely point to a specific disorder. The dataset has of survey like labels that have indirect symptoms framed in questions. In similar manner there are requirements of separate datasets for each individual disorder via which the model predicts the sub type of disorder. SQLite database is required for storing the data of login credentials of registered candidates. Sqlite3 works effectively with flask framework, it provides disk-based database that does not require a separate server process. Many quick web apps nowadays make use of sqlite3 for internal data storage. It is easy to retrieve, store and manipulate user data using sqlite3. Python standard library provides the sqlite3 module

to interact with SQL database without the need of any external installations or setups. Apart from database the system requires various dataset. The system as in the machine learning model requires huge amount of dataset to make accurate predictions. The data set in traditional .csv files are required to train and test the predictive model.

IV. TECHNICAL WORK

In this section, we will discuss our SVM, Random Forest, Naïve Bayes algorithm and Implementation details.

1. SVM ALGORITHM: -

The goal of the SVM method is to discover the best line or decision boundary for categorising n-dimensional space into classes so that subsequent data points can be easily placed in the right category. Hyper-planes are the best choice boundaries. The decision plane (hyper-plane) is a plane that separates items belonging to various classes.

2. Random Forest ALGORITHM: -

The random forest adds extra randomization to the model while growing the trees. Instead than looking for the most important characteristic when dividing a node, it seeks for the best feature from a random group of features. As a result, there is a lot of variation, which makes the model better. Rather than relying on a single decision tree, the random forest gathers forecasts from each tree and predicts the final output based on the majority of votes.

3. Naive Bayes ALGORITHM: -

The naive Bayes classifier is based on Bayes' theorem with the independent assumptions between predictors i.e., it assumes that the presence of a feature in a class is unrelated to any other feature. Even if these characteristics are dependent on one another or on the existence of other characteristics, they all exist independently. Gaussian naive Bayes is a type of naive Bayes that is used to classify data with a binomial (normal) distribution.

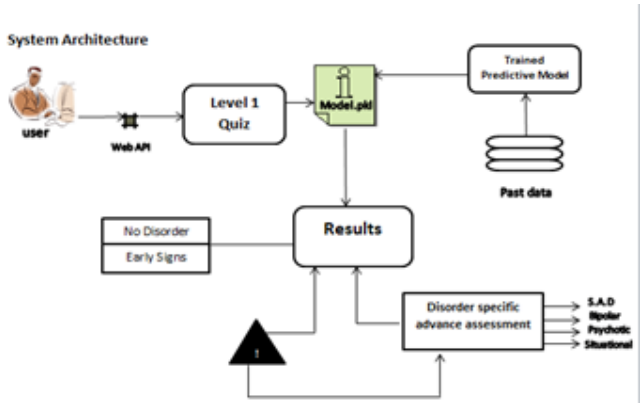
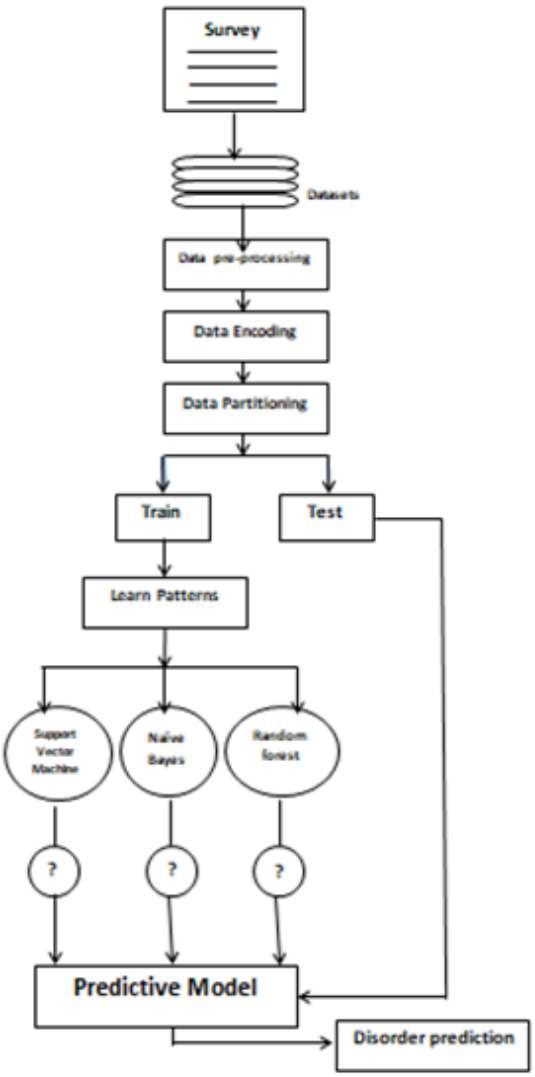


Fig. 2. Workflow of proposed System

4. IMPLEMENTATION: -

This system has multiple datasets involved at every step. The preparatory data acquired is transformed into labels depicting questions that are understood by general users but have the main symptoms lying in them. These labels are answered in yes/no values, which are then converted into binary flags. This we call, as the data pre-processing step in our system. Any model expects data to be reliable and elastic which may be discrete or continuous in nature. It is then subjected to advance integration and processing that involves normalization of data, transformation and encoding the labels into numerical codes, viz. it assigns a numerical digit to each disorder. Example in our system case-{0-'BipolarDepression', 1-'Clinical Depression', 2-'Early signs of depression', 3-'No symptoms', 4-'Psychosis', 5-'Psychotic Depression', 6-'S.A.D', 7-'Situational Depression'}.The processed dataset is then segregated into training data and test data.

- Output - The photographed input is compared to a saved dataset result or one generated by the user. Clinical, psychotic, SAD, situational, bipolar, psychosis are predicted using ml algorithms..



Formula is: -

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Posterior, Likelihood, Prior, Normalizing constant

$$P(B) = \sum_{Y} P(B|A)P(A)$$

Fig.3. Flowchart of prediction

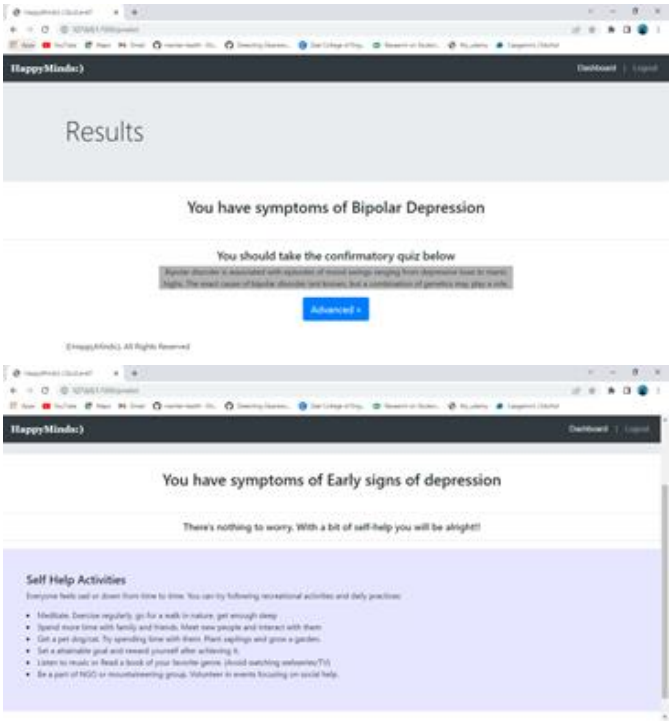


Fig.4. Final prediction and symptoms of disease

By using SVM, Random Forest, Naïve Bayes , we have got a 99.80 percent accuracy of combined model. The training accuracy is shown in Fig.5 and Fig.6.

V. TYPE OF DISEASE TABLE

Disorder	Subtype	Symptoms	Degree of severity
Bipolar	Bipolar I	Oversleeping, Carbohydrates craving	High
Bipolar	Bipolar II	Hypomania,Euthymia	Moderate
SAD	Winter Sad	Concentration issues, Feeling of Worthlessness	Moderate
SAD	Summer Sad	Insomnia,Mood shifts in summer	Moderate
Psychotic	-	Agitation, Anxiety	High
Clinical	-	Sadness, Slowed thinking	Low
Psychosis	-	Delusions, Hallucination	High
Situational	-	Stress, loneliness	High

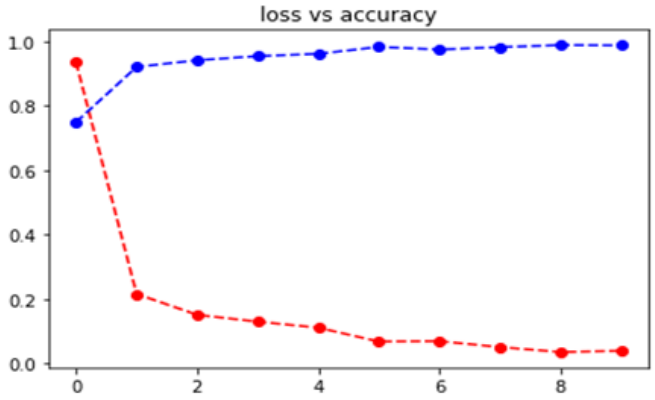


Fig 5. Loss vs accuracy

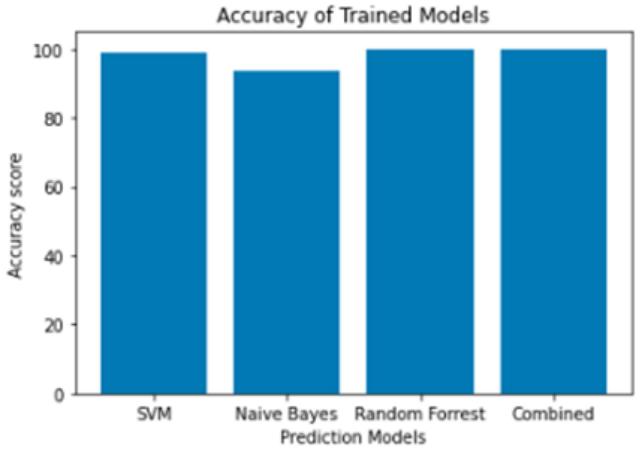


Fig 6. Accuracy of trained model

VI. DISCUSSION AND FUTURE WORK

Data Scientists will continue to work to improve our understanding of mental diseases and extend treatment options using data analysis and machine learning. As more data is collected, collated, and analysed, and algorithms are developed, theories and hypotheses will continue to be tested and proven or disproved. Large sums of money are still being put into projects and studies, such as neuroscience research targeted at understanding the genetics of mental diseases. Medical personnel will be able to better diagnose, treat, and perhaps prevent mental health illnesses as a result of this research. Improved diagnoses will lead to more appropriate treatments for those suffering from mental illness as funding continues to grow. Data Science's position in mental health research will be influenced by technological improvements. Verily, a startup that develops technology to collect health-related data, is urging mental health professionals to promote the use of data science in diagnosis, treatment, and prevention. This would necessitate them relying on evidence-based research and Data Science rather than subjective observations.

The utilisation of this type of big data for psychiatric research certainly has a number of advantages. These data sources, however, have inherent limits, as we have highlighted, and should not be used to substitute

approaches that have already been demonstrated to be useful. Instead, we believe they should work in tandem with randomised control trials, representative surveys, cohort studies, and qualitative investigations, leveraging each's methodological advantages while balancing their limits. We've also discussed how new methodology, such as quasi-experimental designs and embedded RCTs, as well as new recruiting options, can be combined with big data to improve traditional research methods. We are also optimistic that, given how quickly this sector is moving, many more unique applications and methodologies will emerge.

VII. CONCLUSION

Different types of mental illnesses were discussed, as well as reasonable, cost-effective, and practical alternatives for improving mental health care facilities. The digital mental health revolution is currently accelerating beyond the rate of scientific examination, and clinical communities must quickly catch up. Various smart healthcare systems and equipment have been created to minimise the death rate of mentally ill patients and to prevent them from becoming involved in unlawful activities through early detection.

Data science is a fast expanding discipline that has a wide range of applications in mental health research, as we've shown in this post. Most importantly, it allows researchers to do research that incorporates real-world complexity. We believe that the significant advances in mental health research that are beginning to emerge will assist persons with mental illnesses.

VIII. REFERENCES

[1]. T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, "Affective and content analysis of online depression communities.

[2]. Mandar Despande, Vignesh Rao, Depression Detection using emotion artificial . intelligence, 2019, IEEE.

[3]. World health Organization Depression Overview: https://www.who.int/news-room/fact-sheets/detail/depression

[4]. M. J. Friedrich, "Depression is the leading cause of disability around the world," JAMA, vol. 317, no. 15, p. 1517, Apr. 2017.

[5]. S. Paul, S. K. Jandhyala, and T. Basu, "Early detection of signs of anorexia and depression over social media using effective machine learning frameworks," in Proc. CLEF, Aug. 2018, pp. 1–9

[6]. https://www.discoverdatascience.org/social-good/mental-health/