International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

Predictive Models on Early Detection of Mental Health Problems using Big Data and Artifical Intelligence

Dhruvesh Shah[1], Aastha Jain[2]
[1, 2]*B. Tech. Computer Engineering, Mukesh Patel School of Technology Management and Engineering, Mumbai, India*

Abstract: ***In the modern world, due to the increase in mental health disorders and its grave consequences throughout the world, it is of utmost importance that we pay greater attention to mental health and thus find out more about its various disorders and their treatment. As the traditional database management tools cannot handle large datasets, thus using big data analytics tools and techniques is advisable by accumulating the mental health data and further processing it. This paper is putting forward the ideas and strategy of using data mining and prediction models that can be used to efficiently predict the state of the mental health of an individual and thus give out proper remedies for treatment. We put forward a framework that classifies into five kinds of model and discusses predictive modelling in the domain of mental health. We have also proposed a BP Neural Network model for early detection of deteriorating mental health.***
Keywords: Big data, Healthcare, Mental health,Artifical Intelligence ,Neural Networks

I. INTRODUCTION

Around the Globe more than 25% of people in both the developing and developed countries are facing mental health problems. This terabytes and petabytes of data, 4/5th of which is not structured is becoming a strenuous procedure with traditional database management techniques and tools. Introducing Big data Analytics tools and techniques in mental health can greatly improve the quality of treatment. With big data, we can analyze the humongous volume of data that will be produced and further enhances the interpretation of how thoughts, memories, emotions, and actions are comprehended. This can largely help the mental health treatment.

Due to a lack of awareness about mental health and the stigma associated with mental illness along with the limited access to professional help, only 10-12% of these patients will seek assistance. Based on the attitude of people towards mental illness, a study revealed three broad segments of people which are shown in Figure 1.

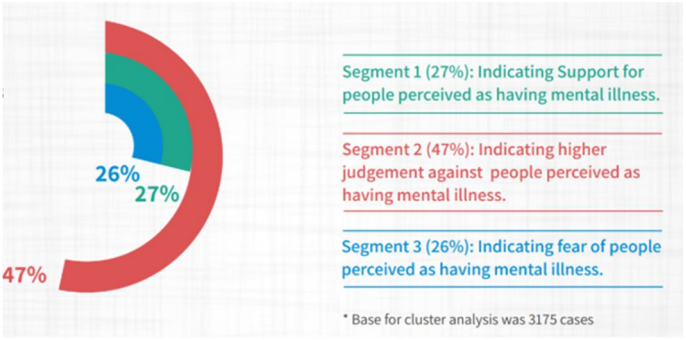

Fig 1. Chart showing percentage of people seeking help

Mental health is a field which needs notable development, as the quantity of mental health data is expanding, our task is to improve the diagnostic and classification analysis to better comprehend the huge volume of data and making this data economical because imprecise and vast volume of data can cause, delays, problems and miscalculation. The tactic is focusing right now on analyzing the different mental disorder and their impact on mental health. Thus, for this firstly the raw data was accumulated, and then by applying the big data tools to extract the data and acquired information from it, a decision was made based on genomic knowledge. This would help making clinical decisions easier. Suitable and correct drug can be given to the patient at the proper time by making genomic analysis. In the research of big data, all varieties of useful apparatus and connections are important tools for collection of information, and all types of tools for accessing information and data mining techniques provide great ease for the research and extraction of extensive scale data.

Principal part of higher education is mental health. However, due to many psychological issues of university students such as, pessimism, self-3abasement, eccentricity, depression, , anxiety and other psychological problems have resurfaced due to negligence of universities in the improvement of students psychological health. And these problems further build up and escalate leading to suicides. To prevent the occurrence of these venomous consequences issued by the psychological traumas, we need a psychological disease predictor to foresee the issues. With the detection of abnormal mental health status of students in time using effective methods and providing psychological help or professional counselling we can save a life.

However, using traditional mental health service agent is extremely difficult to find the abnormal psychological status of students in the universities. Most students tend to hide their mental health problems and are themselves are not conscious of the graveness of their own issues, and thus do not want anyone to know about their own mental health problems because of fear of judgement.

We need to change the focus from the passive traditional mental health services to the active psychological observance. To traverse the psychological status of the students we need to develop a prediction algorithm from the student's daily activity data. Because of big data technology and theory we have found new ideas for research in psychology. When we compare this with the traditional method, big data method is useful and faster. There is extraction of data from the database which then undergoes analysis and processing which has proper objectivity, representation and timeliness.

II. LITERATURE REVIEW

A. Framework

We propose a framework which has major elements which are described below: time (phases), data, decisions, and model types. This framework categorizes and describes the existing applications of predictive modelling in the domain of mental health.

B. Time

Time plays a significant role in predictive modelling. It can be divided into 3 states namely pre-intervention, intervention and post-intervention. Pre-intervention lasts for few days. Intervention lasts for around 6 to 8 weeks and Post-intervention for about 5 months.

C. Data

Data collected at every phase varies. Model 1 requires the least data like the symptoms and personality traits. Model 2 further requires the intervention data and model 3 and 4 consists of long term outcomes and chances of relapse.

D. Decisions

The decisions depend on how sever the patients' needs are. For example, there might be a chance of drop out or relapse if proper guidance and treatment is not given. It is important to keep the intervention in accordance of the client needs and personalizing it according to his or her behaviour.

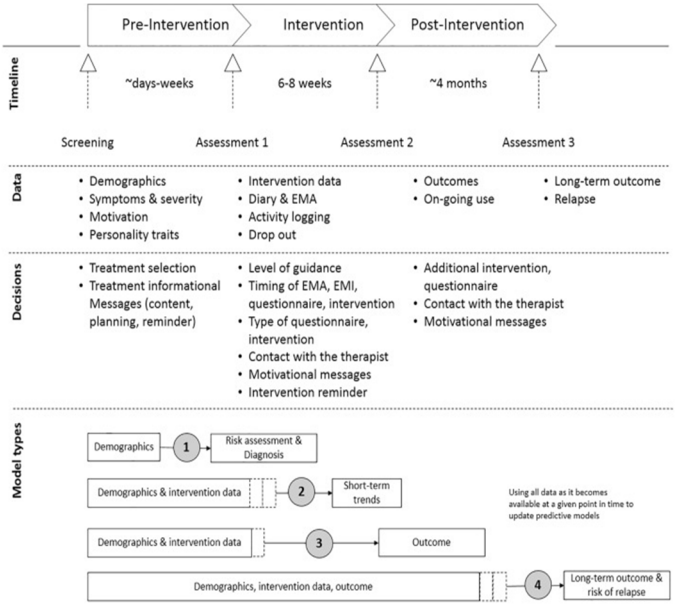

Fig 2. Framework to categorize different predictive models in e-mental health.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

E. Models

1) *Model 1(Risk assessment):* It focuses on predicting the risk of the mental illness and then suggesting the best treatment options available. It makes use of number of variables that are related to socio-demographic characteristics and personality traits. To achieve this goal data mining techniques, have to be used for the detection of risk factors that traditional statistical methods fail to do. Few examples include a supervised modelling technique and logistic regression to determine postpartum depression in women. Other methods include creation of mathematical models for recovery from depression of the general population. Collection of data can be done through social media such as Facebook and twitter. Text mining techniques can be used for the identification of suicide notes. According to studies these predictive models gave better results than mental health professionals. Several studies show that monitoring of these symptoms could be done using predictive models built from smartphone log file data. Factor graph and regression models can be used for the estimation of moods using smartphone measure. Markov models can be used for stress estimation from voice sample. Support vector machines for speech recognition using speech samples. Availability of data in Type 1 models can be through EMA assessments, questionnaires, or shared e-health records. There is a difference in time-span of the data collection which ranges from 1 h ,to days, or weeks. We can improve the predictive accuracy of the models can be increased by taking into account contextual information which can be done using questionnaires or assessments. For example, collection of such data from 27 anxiety clients over thirty days, five times a day . An average accuracy of 84% was achieved from EMA data using Bayesian networks. This model will most likely land into future e-mental health in the form of smartphone applications.
2) *Model 2 (Short- term predictions during on-going treatment):* It targets the patients state during their treatment. This improves treatment adherence and the outcome. It requires weekly administered self report questionnaires. When Type 1 models and applications undergo treatment of short term predictions of health states they are called Type 2 models. Type 2 models are more progressive since data is huge and detailed. . Some examples include oscillating differential equations to predict hypomanic, stable, and depressive episodes in patients diagnosed with bipolar disorder. This model gives insight into medication treatment effects and coupling behavior between patients. Medication treatment not only reduces the increase in mood changes but also lowers the the extent of mood oscillations. Phone usage data and previous EMA measurements can be used for prediction of mood changes and the mood for the next day for a number of students. For phone usage data they include activity levels, app usage ,number of phone calls, and number of messages sent. They apply a variety of data mining techniques including Support Vector Machines, Lasso and Linear regression, and Bayesian Hierarchical Regression.
3) *Model 3(Predicting treatment outcome):* Type 3 models predict the result of the treatment including drop outs. They give us the post-test outcome from the data collected. This has two features: any fluctuation in the health symptoms whether treatment was given as expected. Some of the methods include Hierarchical regression modelling for the approximation of drop-out risk factors and it is noticed that clients with lower levels of self esteem and absence of remedial guidance are more likely to dropout. Logistic regression is used for drop-out estimation of clients with mild panic disorder. Outcome prediction is done using linear regression models and other methods which is based on session-based questionnaires. Treatment resistance is predicted based on self-reported clinical data using naïve Bayes, support vector machine and other methods. Another approach uses free text written by patients .It took out features from the text messages sent by patients undergoing an anxiety disorder to their therapist as part of their treatment. This included words used, opinions and thoughts, feedback and reactions, time for their response, phrases used and length of the emails, and topics that were written in their emails. This proved to be helpful for early detection of clients.
4) *Model 4 (models for deterioration prediction):* It focuses on predicting re-occurrence of the symptoms. This can be soon after the treatment or in the long term. It determines he amount of medical care required. Clinical researchers have used traditional techniques such as logistic regression to anticipated relapse in patients diagnosed with a variety of disorders such as depression, bipolar depression and alcohol misuse. Some of the examples include the outcome of bipolar clients at one year follow up is predicted using clinical data. Based on previous drinking habits, relapse of drinking after 3 to 6 months is predicted using logistic regression. Mood, social ad cognitive vulnerability helps in estimation of depression using regression trees. These models will find their path into mobile applications and will most likely perform better because the models will then be able to take patient's current contextual information into account, resulting in various options to personalize interventions and give the user correct assumptions and treatment for the same.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

III. TEXT CLASSIFICATION TECHNIQUES

A. K-Nearest Neighbors

The KNN algorithm assumes that alike things always are near each other. In other words, alike things are close to each other. KNN algorithm is dependent on the idea of similarity. Similarity is sometimes distance, or closeness. It can also be proximity. Depending on the problem we are solving, there are various preferred ways of calculating distance. However, the the Euclidean distance or straight line is a popular choice.

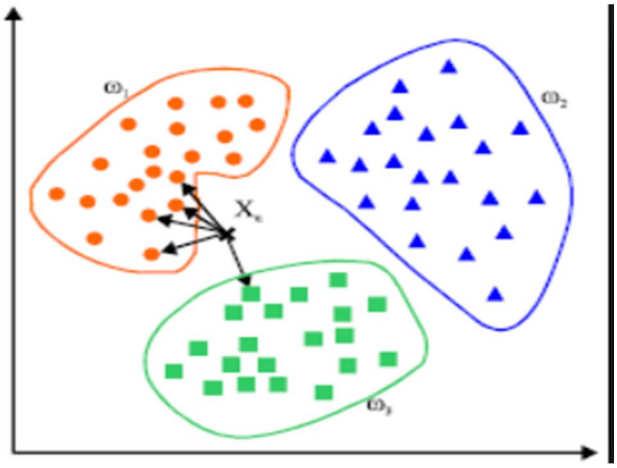

Fig 3. k Nearest Neighbor technique

B. Support Vector Machine

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which is useful for classification as well as regression challenges. It is more commonly used in classification problems. In this algorithm, each data item is plotted as a point in n-dimensional space (where n is number of features) with the value of each feature being the value of a particular coordinate. Thus, by finding the hyper-plane that differentiate the two classes we then accomplish the classification.

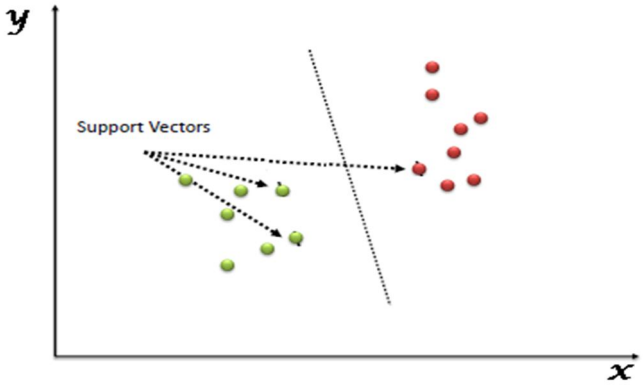

Fig 4.Support Vector Machine technique

C. Naive Bayes Classifier

Naive Bayes is a classification algorithm which is further divided for binary (two-class) and multi-class classification problems. Binary or categorical input values make this algorithm quite easy to understand.

It is also known as naive Bayes *or* idiot Bayes. This is due to the reason that to make the calculation tractable, the calculation of the probabilities for each hypothesis are simplified. Instead of calculating the values of each attribute value P(d1, d2, d3|h), they are presumed to be conditionally independent given the target value and calculated as P(d1|h) * P(d2|H) and so on.

The approach that the attributes do no interact is a very bold assumption but performs surprisingly well on data where this assumption does not hold

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

IV. BOOK CLASSIFIER BASED ON TEXT CLASSIFICATION TECHNOLOGY:

Text classification is used in many applications like web page classification, classic spam filtering and news classification. We try to find a correlation between the reading habits of patients and their depressive tendency based on the data collected and the questionnaire results. We collect book records of all the patients and find a relation to their depressive tendency. Using different text classification algorithms namely kNN, SVM and native Bayes algorithm. They have different mathematical principles and their features are different too. We will try all the methods and choose the most appropriate method according to the performance and its efficiency. In relation to accuracy, Bayes and SVM classification have maximum accuracy of about 0.823. Coming to time and cost Naïve Bayesian has the least whereas kNN has the maximum. Naïve Bayesian uses simple addition and subtraction and hence consumes lesser time compared to SVM and kNN which uses multiplication, division and square root operations. Therefore, Naïve Bayesian is the best method and with increase in sample data set the accuracy can be increased further.

Algorithm	1000	2000	3000	4000	5000	5500
kNN	0.604	0.596	0.693	0.685	0.709	0.766
SVM	0.668	0.742	0.768	0.796	0.780	0.82
Bayes	0.452	0.512	0.614	0.612	0.632	0.642
Multi-NB	0.704	0.724	0.790	0.803	0.814	0.823
Bernoulli-NB	0.524	0.574	0.597	0.643	0.657	0.654

Fig 3. Accuracy of classification algorithms using different sample sizes.

Algorithm	1000	2000	3000	4000	5000	5500
kNN	5.7	19.6	40.5	65.1	100.1	150.5
SVM	3.1	14.4	29.2	54.9	96.8	120.4
Bayes	1.5	2.6	5.2	6.7	9.7	12.1
Multi-NB	0.9	1.7	2.5	3.7	6.8	9.9
Bernoulli-NB	0.9	1.7	2.8	4.2	6.6	9.8

Fig 4. Accuracy of classification algorithms using different sample sizes.

V. PSYCHOLOGICAL CRISIS PREVENTION SYSTEM BASED ON BP NEURAL NETWORK

The system primary task is to collect information about patients, perform psychological health level survey, set up patient psychological archives, statistics and analysis of their psychological services is also done. The system has an exclusive management capacity of psychological measurement scale, prevailing statistics, analysis, screening, appropriate query and print function

The information communication data interaction between each model is smooth, faster system response speed, user interaction interface is approachable, and operation is easy. The embedded data mining module has good compatibility with the authentic system, data exchange is well, self-protection of the embedded module and reliability requirements is high. It can implement a more complete data mining model, which can satisfy mining result comparison of different data mining model. The system is divided into data layer, data mining layer and user interface layer. The data layer is largely used to store patients psychological raw data. Data mining layer carries out data mining for target data set through the data mining model. User interface layer can view the data mining classification outcomes, evaluate and analyze the mining results and mine the knowledge rules. The data is stored in the psychological census database. After data preprocessing, we obtain mining data set. BP neural network is used for data mining. The final result is provided for mental workers

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

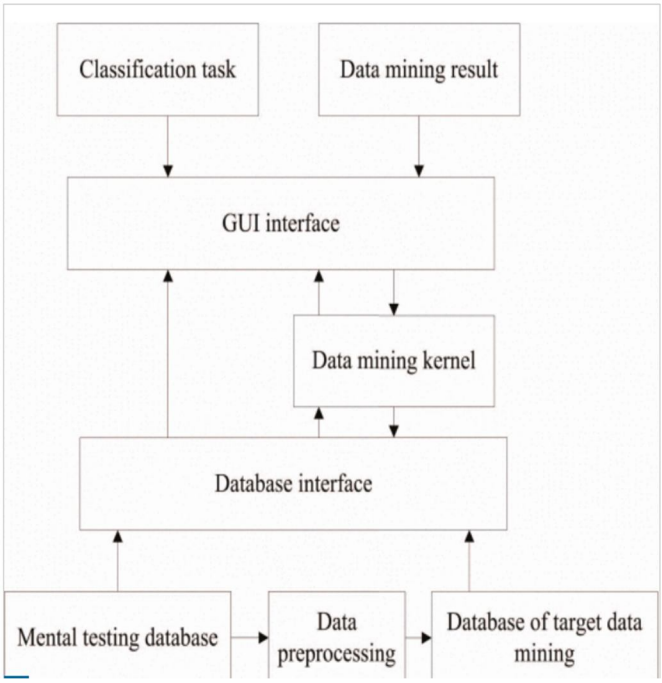

Fig 5. Psychological crisis prevention system structure

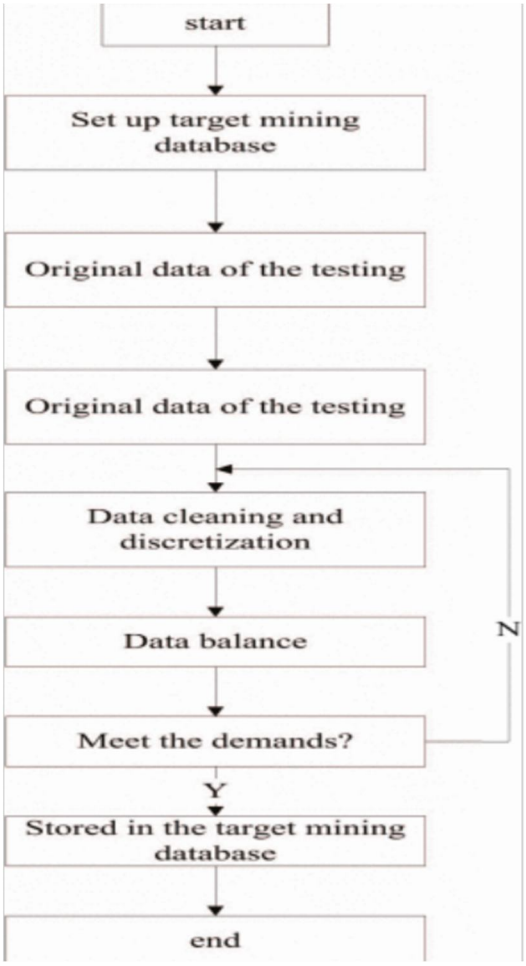

Fig 6. Preprocessing database

VI. INFERENCE

We provided an introduction to the different predictive models and their application in mental health. This provides a mutual ground between data mining community and the researchers. Our review proposes that mental health problems research should be focused on these model calculations rather than the old school traditional models. There is also possibility for further development in the applications of predictive modelling. We realized that numerous reports do not use self-governing test sets to assess new modelling techniques. Appropriate testing of models is very important to come down to a generalized model. We first provided an outline of the data mining field and the techniques used for predictive modeling. Later, we characterized predictive modeling research in mental health care on three aspects: Firstly the time which differs with the treatment , then comes types of data available(e.g., questionnaire data, smartphone data) and lastly the type of clinical choice (these decisions include treatment personalization and selection). Focusing on these three aspects , we introduced an outline that makes sure that these four model types can be used to categorize current and upcoming applications. After discussing the four models we can infer that Model 1 is best for early risk assessment and diagnosis. Model 2 identifies risk of drop-out and is used for short-term trends. Model 3 is used to predict the therapy outcomes and finally Model 4 makes sure that the results are stabilized and prevents relapse.

VII. CONCLUSION

Thus, we can conclude that different predictive models are suitable according to different conditions. We can further conclude that seeing the number of people suffering from mental health in today's world, it is of utmost necessity to treat it professionally and meticulously. Due to the vast amount of data, it is necessary that big data techniques and models should be introduced instead of the traditional approach. In the previous section we studied the various techniques and concluded the Bayesian to be the best one. Therefore, in conclusion we need to increase awareness about the better treatments available due to early prediction of mental health issues, so that more people step forward and break the stigma surrounding mental health.

REFERENCES

[1] Yuji Hou, Jingjing Xu, Yixin Huang, Xiofeng, " A Big Data Application to Predict Depression in the University Based on the Reading habits", in the 3rd International Conference on Systems and Informatics (ICSAI), 2016.

[2] Jian Quihua, "Data Mining and Management System in Design and Application for College Student Mental Health", in 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS).

[3] Dennis Becker,Ward van breda,Mark Hoogendoorn, "Predictive Modelling in E-mental Health : A Common Language Framework",Elsevier Internet Inventions 2018.

[4] Priyanka Dhaka,Rahul Johari, "Big Data Application :Study and Archival of Mental Health Data,using MongoDB" , in Internation Conference on Electrical , Electronic and optimization techniques(ICEEOT),2016.