MENTAL ILLNESS PREDICTION USING MACHINE LEARNING

Swathi, S.[1]

Assistant Professor, Department of Computer
Science and Engineering,
Sri Sai Ram Institute of Technology,
(laxmiswathiks@gmail.com)

Kamali P[2]

Student, Department of Computer Science and
Engineering,
Sri Sai Ram Institute of Technology,
(sit20cs105@sairamtap.edu.in)

Gowrishankari M[3]

Student, Department of Computer Science and
Engineering,
Sri Sai Ram Institute of Technology,
(sit20cs083@sairamtap.edu.in)

Archanaa A[4]

Student, Department of Computer Science and
Engineering,
Sri Sai Ram Institute of Technology,
(sit20cs078@sairamtap.edu.in)

Abstract

An automated approach for detecting mental health is being developed in this project. In order to detect mental health issues including stress, anxiety, depression, and other issues, we are working to build an end-to-end method. When a mental health issue is detected, our system will suggest the appropriate therapies. If the issue is severe, the system will assist in contacting the appropriate psychiatrists or doctors. To do this, we will use a variety of machine learning (ML) techniques and classifiers to create a prediction model on historical data, plot the results, and determine the model's accuracy using test data. We need to deploy the model online in order to utilize it to predict the fresh data. Then, using Flask, we will build a web application on it.

Keywords*: Machine learning algorithms, Classifiers, Random forest, KNN, Exploratory Data Analysis, Data Visualization*

1. INTRODUCTION

The need for high-quality medical care and the prevalence of mental health problems have sparked research into possible applications of machine learning to these problems[1]. A recent analysis of machine learning techniques for foretelling mental health disorders is provided in this publication. Mental disease diagnoses are frequently made based on the patient's self-report, which necessitates the use of questionnaires designed to pinpoint specific emotional or social interaction patterns. [7]. With proper care and therapy, many people with mental illness or an emotional disorder should be able to recover [9].

The most common mental health issues include worry, stress, and sleep disorders. Millions of people both in India and around the world are impacted by these challenges. Unfortunately, a lot of people who are struggling with one or more of these challenges occasionally feel like they are fighting alone, and this can have serious consequences, like suicide. The users of this system may be able to resolve these issues. We provide a critical study of machine learning approaches used to detect mental wellness.

We use machine learning techniques to ascertain a user's mental state utilizing a combination of self-provided data and passive data from our website after training the platform. Users' information is gathered, and then it is usedto provide therapies.

The goal of machine learning is to create systems that can learn from experience by utilizing sophisticated statistical and probabilistic methods. It is thought to be a very valuable tool for making mental health predictions. It is enabling several researchers to gather crucial data and create tailored experiences, automated intelligent systems, and other things [4]. Future events have been predicted and categorized using the commonly used machine learning algorithms support vector machines, random forests, and artificial neural networks [5]. Machine learning is techniques that aim to construct systems that can improve through give a critical analysis of mental health detection using machine learning techniques experience by using advanced statistical and probabilistic techniques. It is believed to be a significantly useful tool to help in predicting mental health. To help collect accurate and useful information, referrals from subject-matter experts inmental health are also necessary [12].

It is allowing many researchers to acquire important information from the data, provide personalized experiences, and develop automated intelligent systems [4]. The widely used algorithms in the field of machine learning such as support vector machine, random forest, and artificial neural networks have been utilized to forecast and categorize the future events [5].

2. PROBLEM STATEMENT

The rate of mental health issues and psychological issues has increased dramatically as a result of the current societal changes in the social environment. According to the WHO, 450 million people worldwide suffer from mental illness. Undoubtedly, a person's emotions, intellect, and ability to communicate with others are all affected by mental illness, which is a health issue. These problems have demonstrated that mental illness has major societal repercussions and necessitates innovative techniques for prevention. And also traditional machine learning faces common training problems, such as overfitting, model interpretation, and generalization [15]. As a result, the researcher turned to deep learning techniques, which have been an effective tool in recent years. This is because machine learning can solve more complex problems, especially in health data [15], [16].

3. LITERATURE SURVEY

Stress, depression, and other psychological health conditions have grown quite widespread among the general public in today's fast-paced environment. In this study, machine learning algorithms were used to forecast levels of stress, anxiety, and depression.

Data were gathered from employed and unemployed people across many cultures and communities using the Depression, Anxiety and Stress Scale questionnaire in order to apply these algorithms (DASS 21). Five distinct machine learning algorithms were used to predict the occurrence of anxiety, sadness, and stress on five different severity levels. Because these algorithms are extremely accurate, they are well suited to forecasting psychological issues. In order to help choose the Random Forest classifier as the highest accuracy model among the five applied algorithms, the f1 score metric was included. The specificity parameter also showed that the algorithms were quite sensitive to negative findings.

The use of social media in the context of eHealth has significantly increased during the past ten years. Only a few research have used machine learning (ML) algorithms to categories the mental diseases of patients, including schizophrenia, autism, obsessive-compulsive

accurately categories data into four types of mental disorders by utilizing the capabilities of a popular classifier called "XGBoost" (Schizophrenia, Autism, OCD and PTSD). The experimental finding shows that, when compared to state-of-the-art classifiers, the suggested methodology is more effective in classifying patient data. The proposed model was effective, as evidenced by the 68% accuracy that was attained.

For early intervention and to stop serious negative consequences in the future, it is essential to predict which kids will go on to develop mental health symptoms as teenagers. Our goal was to create a model that might anticipate mental health issues in mid- adolescence. Look at whether or not logistic regression will perform better than machine learning approaches like random forest, support vector machines, neural networks, and XGBoost.

The results of the Strengths and Difficulties Questionnaire indicated mental health difficulties. Despite some variation in model performance, the random forest model demonstrated non-significant superiority (AUC = 0.739, 95% CI 0.708-0.769), which was followed closely by support vector machines (AUC

= 0.735, 95% CI 0.707-0.764). Future investigations should use parent-rated assessments whenever possible.

4. EXISTING MODEL

These are the literature reviews that were used to develop the current system. Each of the systems has drawbacks that can be categorized as follows.

For categorization, Random Forest makes use of both structured and unstructured data. Supervised learning, in more specific terms, is a classification method that makes use of structured training data [11]. Other machine learning algorithms, which only deal with structured data and take a long time to compute, are sluggish because they store all of the data as a training dataset and employ complicated calculation techniques. Despite being an effective technology, it is only used in hospitals and does not capture real-time patient data to monitor patients as they receive treatment. Hence, it is crucial for researchers to explore and analyze the data using different machine learning algorithms in order to select the one with the highest accuracy[12]. Accuracy, the effectiveness of each classifier using all machine learning techniques is compared [14].

By reviewing the algorithms and parameters used in each system, this review paper analyses various systems for monitoring mental health, including virtual counseling, precision therapy, and diagnostic systems. It then proposes a system that combines the systems mentioned above and is intended to provide

individualized mental care. This review piece is onlyused for academic purposes.

5. MODEL PROPOSED

We incorporate KNN classifiers in our ML model.In comparison to other models already in use, ourmodel will provide accurate predictions faster (Figure 1).
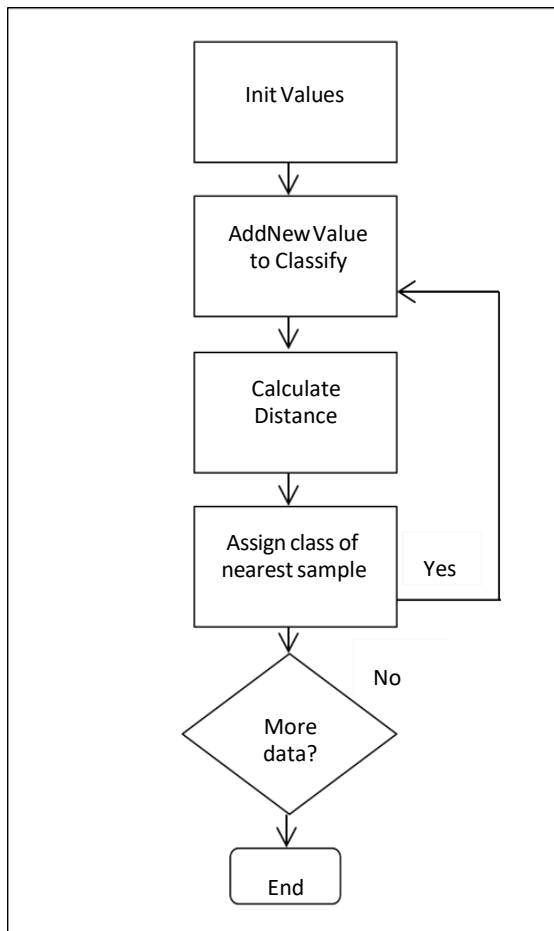
Figure 1 - KNN classifiers

The KNN method makes the assumption that the case/data and existing cases are comparable and places the new case in the category that best matches the existing categories. Regression and classification do not require any training time. The current models are not very user-friendly and are solely used for prediction. However, our technology will make predictions and offer treatments based on the level of prediction. So, the user will find it simple to utilize. We must create a website in order to collect data from the user on the front end. We will employ computer languages to create a website. After that, the user will login using their personal data. They should then take the test after that. Based on that evaluation, we gather information from users to forecast mental disease.

For many businesses, creating machine learning models is a new activity that might be intimidating. Creating an AI model requires effort, experimentation, and creativity, even for people with prior machine learning knowledge. However, the procedure for creating data-centric projects is well known. In order to develop predictions and accomplish its objective, a machine learning model must first learn from training data, generalize what it has learned, and then apply
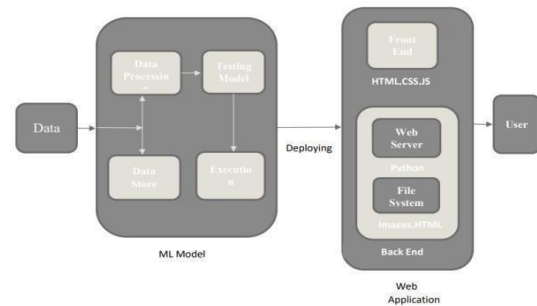
Figure 2 - Implementation

that knowledge to fresh, unexplored data. Information gathering is the next step. Train the model to take in knowledge from the superior data you has gathered using a range of techniques and algorithms.

Depending on the data samples obtained and the data's properties, the machine learning models' performance will change. Additionally, preparatory tasks like data cleaning and parameter tuning might have an impact on machine learning models in order to get the best outcomes. Therefore, in order to choose the machine learning algorithm with the best accuracy, it is essential for researchers to investigate and analyze the data using various machine learning algorithms [12]. This step calls for the selection of a method, hyper parameter configuration and adjustment, ensemble model building and testing, model validation, and model optimization. The machine learning algorithm receives the data, which is then divided into two portions for training and testing. After that, the model's performance will be evaluated. Model evaluation can be compared to machine learning quality control.

The effectiveness of the evaluation against metrics and requirements will decide how the model performs in the real world. It examines the machine learning model's performance before making a prediction about the result. Just because you're running the model and monitoring its performance doesn't mean you're done. The maxim "start small, think big, iterate often" is frequently cited as the secret to successful technological deployment. What type of mental

disease they have, how they will be treated, and the capability of continuously evaluating and monitoring the model's effectiveness. Finally, we will launch the created web application.

Hardware Requirements

- Minimum 8GB RAM
- 16GB RAM and above is recommended for mostdeep learning tasks
- When it comes to CPU, a minimum of 7th generation (Intel Core i7 processor) is recommended.
- Getting Intel Core i5 with Turbo Boosts.

Software Requirements

Google Research's Collaboratory, or "Collab" for short, is a product which is a web-based Python editor that allows anyone to write and run arbitrary Python code. It's notably useful for machine learning, data analysis, and education.

5. ML MODEL IMPLEMENTATION

Classification strategies were employed in earlier research [17], [18], [19], [20]-[28], [29], and [34] to
identify a variety of mental health issues, including stress, suicide ideation, depression, and distress.

i. Importing the data from csv files

A frequently used function for importing datasets may be found in the pandas package. It enables the import of data in several formats, including csv and xlsx files. The same function will be used. Read the data after importing the pandas package, which includes all data operations done before building a machine learning model. We will now investigate the encoder data set.

ii. Data Visualization

It is a graphical representation of the data that is used, among other things, to look for patterns, outliers, and data distribution. Python supports the data visualization tools Matplotlib, Seaborn, and others. We will make use of the seaborn library to visualize the pair plots. Use the code below to verify the pair plot. Once the Seaborn library has been imported, the pair plot will be printed.

iii. Splitting data into training and testing

Before any data is used to build a machine learning model, it is always divided into two categories: training and testing. We never allow testing data to be exposed and only ever give the model training data for training. After the model has been trained using the testing data, we use it to make predictions over that data, which is stored in a single variable called y pred. It can also be kept in a different variable. We will start by defining the independent and dependent variables, X and y, respectively.

iv. Building Machine Learning Models

Now, Random Forest and KNN, two different machine learning algorithms, will be used to construct the machine learning model. Logistic regression is based on linear models, whereas random forest is an ensemble method. They will be imported first, and both models will then get training data. Once it has been trained, we will compute predictions over testing data and store them in separate variables.

v. Prediction by the Models

Exploratory data analysis (EDA) is a technique for getting acquainted with a set of data. We ascertain the data's shape, type, and any missing values. Now, we will compute predictions for a few rows to check the model's predictive accuracy.

vi. Model Evaluation

Model evaluation is a method for assessing the effectiveness of the model by computing different error measures. Performance in classification and regression tasks is evaluated using a variety of error metrics, including accuracy, confusion matrix, mean squared error, and mean absolute error. We built our classification model; thus we calculated metrics to evaluate the classification model. The confusion matrix and classification report will be computed after the accuracy score.

6. CONCLUSION AND FUTURE ENHANCEMENT

This article introduces a web tool for monitoring a person's mental health online. As a result, clients are able to live stress-free. In today's hectic world, everything has become complicated. The goal of our project is to maintain a simple, calm way of life for each individual's mental health. Customers can easily use our project because it is being released as a web application.

In the future, we're going to provide chat bot functionality as well. Now, chat bots only have text- based systems, but in the future, we will integrate a voice recognition system. Therefore, based on the severity of the mental illness We provide activities and therapies including music, stories with motivation, meditation, and straightforward tasks.

7. REFERENCES

[1] Discovery of Mental Wellness via Social Analytic for Livability in an Urban City.
Author:Kar Way Tan- Published in 2021

[2] Analytic Vidhya,12018,[online]Available: https://www.analyticsvidhya.com/blog/2017/09/unde rstaing-supportvector-machine-examplecode/

[3] Danilo Bzdok and Andreas Meyer- Lindenberg, "Machine learning for precision psychiatry", Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, vol. 3, no. 3, pp. 223-230, March2018.

[4] A Lanata, G Valenza, M Nardelli, C Gentili and EP Scilingo, "Complexity Index from a Personalized Wearable Monitoring System for Assessing Remission in Mental Health", IEEE J Biomed Health Inform, Sep 29.- Published in 2015

[5] Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges.Author: Jetli Chung1and Jason Teo, Faculty of Computing and Informatics, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah,Malaysia available at https://doi.org/10.1155/2022/9970363-published in Jan 2022

[6] G. Miner, L. Miner, M. Goldstein et al., Practical Predictive Analytics and Decisioning Systems for Medicine: Informatics Accuracy and Cost-

Effectiveness for Healthcare Administration and Delivery Including Medical Research, Academic Press, Inc., Cambridge, MA, USA, 1st edition, 2016.

[7] M. Hamilton, "Development of a rating scale for primary depressive illness," British Journal of Social and Clinical Psychology, vol. 6, no. 4, pp. 278–296, 1967. View at: • Publisher Site | • Google Scholar-published in 1967.

[8] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, American Psychiatric Association, Reston, VA, USA, 2013.

[9] M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," Data Science. Kai Zhao, Hon-Cheong So, vol. 349,no. 6245, pp. 255–260, 2015. View at: Publisher Site | Google Scholar, published in 2015

[10]. E. G. Pintelas, T. Kotsilieris, I. E. Livieris, and P. Pintelas, "A review of machine learning prediction methods for anxiety disorders," in Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info- Exclusion—DSAI 2018, Thessaloniki, Greece, 2018.

[11]. Application of Machine Learning Methods in Mental Health Detection: A Systematic Review.
Authors: Rohizah Abd Rahman, Khairuddin Omar, Shahrul Azman Mohd Noah, Mohd Shahrul Nizam Mohd Danuri, Mohammed Ali Al-Garadi.
Publisher:IEEE-published in October 2020

[12]. W. Bosl, A. Tierney, H. Tager-Flusberg, and C. Nelson, "EEG complexity as a biomarker for autism spectrum disorder risk," BMC Medicine, vol. 9, no. 1, 2011. View at: Publisher Site | GoogleScholar

[13]. H. Lin, J. Jia, Q. Guo, Y. Xue, Q. Li, J. Huang, et al., "User-level psychological stress detection from social media using deep neural network", Proc. ACM Int. Conf. Multimedia (MM), pp. 507-516, 2014.

[14]. D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, et al., "Deep learning for health informatics", IEEE J. Biomed. Health Informat., vol. 21, no. 1, pp. 4-21, Jan. 2017.

[15]. X. Li, Y. Jiang, M. Li and S. Yin, "Lightweight attention convolutional neural network for retinal

vessel segmentation", IEEE Trans. Ind. Informat., May 2020.

[16]. H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, et al., "Detecting stress based on social interactions in social networks", IEEE Trans. Knowl. Data Eng., vol. 29, no. 9, pp. 1820-1833, Sep. 2017.

[17]. M. Kandias, D. Gritzalis, V. Stavrou and K. Nikoloulis, "Stress level detection via OSN usage pattern and chronicity analysis: An OSINT threat intelligence module", Comput. Secur., vol. 69, pp. 3-17, Aug. 2017.

[18]. M. Thelwall, "TensiStrength: Stress and relaxation magnitude detection for social media texts", Inf. Process. Manage., vol. 53, no. 1, pp. 106- 121, Jan. 2017.

[19]. B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris and H. Christensen, "Detecting suicidality on Twitter", Internet Intervent., vol. 2, no. 2, pp. 183- 188, May 2015.

[20]. H. Lin, J. Jia, Q. Guo, Y. Xue, J. Huang, L. Cai, et al., "Psychological stress detection from cross- media microblog data using deep sparse neural network", Proc. IEEE Int. Conf. Multimedia Expo (ICME), pp. 1-6, Jul. 2014.

[21]. S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh and H. Ohsaki, "Recognizing depression from Twitter activity", Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. (CHI), pp. 3187-3196, 2015.

[22]. S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh and H. Ohsaki, "Recognizing depression from Twitter activity", Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. (CHI), pp. 3187-3196, 2015.

[23]. X. Huang, L. Zhang, D. Chiu, T. Liu, X. Li and T. Zhu, "Detecting suicidal ideation in Chinese microblogs with psychological lexicons", Proc. IEEE 11th Int. Conf. Ubiquitous Intell. Comput. IEEE 11th Int. Conf. Autonomic Trusted Comput. IEEE 14th Int. Conf. Scalable Comput. Commun. Associated Workshops, pp. 844-849, Dec. 2014.

[24]. M. De Choudhury, M. Gamon, S. Counts and E. Horvitz, "Predicting depression via social media", Proc. 7th Int. AAAI Conf. Weblogs Social Media (ICWSM), pp. 128-137, 2013.

[25]. M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence", Proc. Int. Conf. Intell. Sustain. Syst. (ICISS), pp. 858-862, Dec. 2017.

[26]. X. Wang, C. Zhang and L. Sun, "An improved model for depression detection in micro-blog social network", Proc. IEEE 13th Int. Conf. Data Mining Workshops, pp. 80-87, Dec. 2013.

[27]. Y. Xue, Q. Li, L. Jin, L. Feng, D. A. Clifton and G. D. Clifford, "Detecting adolescent psychological pressures from micro-blog", Proc. Int. Conf. Health Inf. Sci., pp. 83-94, 2014.

[28]. A. Wongkoblap, M. A. Vadillo and V. Curcin, "A multilevel predictive model for detecting social network users with depression", Proc. IEEE Int. Conf. Healthcare Informat. (ICHI), pp. 130-135, Jun. 2018.

[29]. E. Saravia, C.-H. Chang, R. J. De Lorenzo and Y.-S. Chen, "MIDAS: Mental illness detection and analysis via social media", Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), pp. 1418-1421, Aug. 2016.