



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 8, August 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Prediction of Mental Health (Depression) Using Data Science Technique

B Soumya, Arya Pandey, Sridhar S

Student, Dept. of CSE, SRM Institute of Science and Technology, Chennai, India

Student, Dept. of CSE, SRM Institute of Science and Technology, Chennai, India

Assistant Professor, Dept. of CSE, SRM Institute of Science and Technology, Chennai, India

ABSTRACT: In recent times, an individual's mental health has become a game-changer in his or her life. It has a repercussion on each and every decision a person makes. Depression is amongst the most leading causes of mental disability in the world. It is a medical ailment that limits your ability to function and disrupts your mood has been identified as one of the most debilitating conditions for people, families, and society. Timely Surveillance of this particular mental health disorder allows specialists to treat them more effectively and enhances patients' quality of life. The procedure of diagnosing mental illness entails numerous phases and is not simple. As a result, there is a critical need to address common mental health concerns that impact everyone and can lead to more serious problems if not identified earlier. Gender, Marital status, Age, Education, Total expense and Income level have all been recognized as essential socio demographic variables in determining variations in depression occurrence. The proposed model employs the Random Forest classifier algorithm to predict whether or not a person is depressed, followed by a prepared questionnaire that will notify the person of the severity of depression based on the depression score calculated using the responses provided by the user in the questionnaire, and subsequently the remedies

KEYWORDS: Depression, Random Forest Classifier, Machine Learning, Data Science

I. INTRODUCTION

“The emotional, social, and psychological wellness of a person is referred to as mental health.” A person’s “cognitive” ability is affected by it. It has now evolved into a game-changing factor in anyone's life. In our conservative society, mental health talk is usually regarded as rare, if not taboo, which has made it even more vital to be openly discussed. It's always a good idea to have a proper and healthy conversation about mental health. Maintenance of positive mental health allows people to reach their maximum strengths, which increases productivity in the long run. Depression is a common disorder marked by an unrelenting sad mood or waning interest in activities that disrupts a person's normal daily routine and causes suffering. An amalgamation of biological, physical, or social changes caused by grief could be one of the possible explanations. The following are some of the most common symptoms:-

- Isolation from day-to-day events and community interactions
- Intake of drugs
- Mood fluctuations that are too much to bear
- Self-Harm

Depression is treated with a combination of medication and verbal therapy sessions under the supervision of a trained psychologist. When a person shows indications of depression, they should schedule an appointment with a psychologist to address the issue. Lower objective socioeconomic status (SES)—as defined by socio-demographic characteristics—has been related to an increased risk of depression. On the other hand, the link between social status perceptions and depression instances is less well recognized. If emergency assistance from a psychiatrist is necessary, this information will be provided. The project is focused on predicting depression in a person using random forest classifier algorithm based on socio demographic features such as age, marriage status, education, family members, total income, and total expense, then to evaluate the level of depression (Mild Depression, Moderate Depression, and Severe Depression), and to generate a depression score based on the user's answers. Treatments will be recommended after the degree of depression has been evaluated, and a form will be issued in case emergency assistance from a psychotherapist is necessary.



II. RELATED WORK

1. “Detecting Mental Disorders in Social Media Through Emotional Patterns - The case of Anorexia and Depression”. In this paper, a combination of both proposed models (BoSE and Δ -BoSE representations) was used using the late fusion approach and a F1 Score of 0.64 was obtained. The proposed model was evaluated against “ERisk2018” datasets which contains several social media posts from Reddit Platform to detect Depression and Anorexia [1]
2. “Predicting the Utilization of Mental Health Treatment with Various Machine Learning Algorithms”. The fundamental point of the paper is that to conclude, a machine learning algorithm is used to predict the individual’s feasibility being treated for mental health issues with a “high accuracy of around 76.3% to 82.5%”. [4]
3. “Prediction of Mental Health Problems Among Children Using Machine Learning Techniques”. The study evaluated the performance of 8 machine learning algorithms. The results show that 3 algorithms, “LAD Tree”, “Multilayer Perceptron”, and “Multiclass Classifier” have more precise outcomes than the others. The data set is found to be fairly small, hence the findings can be extended to a larger data set in the future to improve the accuracy score. Prior to using any technique in real-world prediction, classifiers must be trained. [2]
4. “Machine Learning-based Approach for Depression Detection in Twitter Using Content and Activity Features”. A binary classification problem is defined in this study as determining whether or not based upon his tweets and Twitter profile activity, a person is depressed. With application of synonym features in addition to the training and testing sets, SVM-L model outperforms other models with an accuracy of 82.5%. [3]
5. “Predicting Individuals Mental Health Status in Kenya using Machine Learning Methods”. In this paper, 17 different algorithms were compared on the basis of performance metrics, out of which SVM_RDF, SVM_ploy, ADA and Voting model had the best performing accuracy and the same accuracy value of 85.13% [7]
6. “Depression Detection Based on Deep Distribution Learning”. In this paper, a deep distribution learning architecture approach is introduced for detecting depression. An expected value of the estimated depression distribution provides the projected value. An expected loss is used to train the model in order to estimate the distribution of data over depression levels. [8]

III. PROPOSED SYSTEM

This section will talk about our proposed methodology which consists of modules of data preprocessing, exploratory data analysis, implementation of algorithm, deployment of the model and the final demonstrative output. Figure 1 shows the overall workflow of the proposed system.

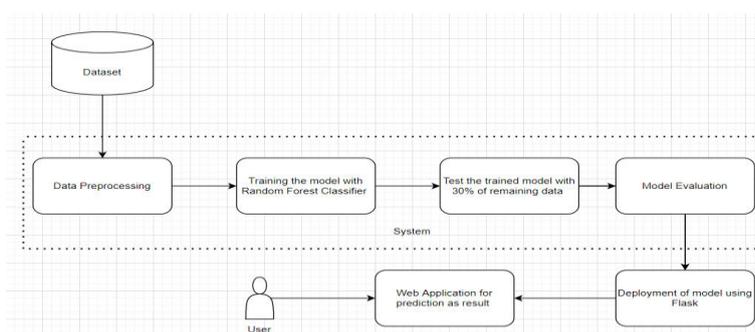


Figure 1. Architecture Diagram of the Project Workflow

A) Modules:

The proposed system consists of five modules.

1. Data Collection

We have taken the dataset from a social survey conducted in the rural area of Kenya, Africa.

2. Data Preprocessing

The data that was gathered might include missing values, resulting in dataset inconsistency. To get better results, data



should be preprocessed to boost the algorithm's efficiency. To clean data EDA(exploratory data analysis) to apprehend information like variable identification, missing value treatments and analyze the data validation .The outliers have to be removed using Z-score where data points whose Z-score falls out of 3rd deviation is detected and the data and the data is removed. In order to understand the patterns, trends and outliers better in the data set. Data visualization is being used.

We have analyzed Massive amounts of information using Univariate, Bivariate, Multivariate Analysis and making data-driven decisions. In these module we have also implemented a Univariate feature selection method called K-Best selection where scikit-learn library provides the SelectKBest class that can be determined to pick a given number of characteristics, it is employed with a suite of different statistical tests. The top 7 features are selected using this selection method namely age, gender, education levels, marital status, family members , living expenses and total income.

3. Implementation of Model

In this module, we will split our dataset using the “scikit -learn” split train-test function. The train test split is 80-20 and we initialize the random state to 1. The algorithm is trained with the training data and the model obtained is tested on the training data. The random forest classifier algorithm is called using the “sklearn.ensemble” library.

4. Model Evaluation

The model is evaluated using a confusion matrix and performance metrics like precision, recall, accuracy, sensitivity, specificity values are calculated using a set of formulae. An accuracy of 93.28%, precision of 92%, sensitivity of 84%, F1 score of 92%, and recall of 92% is obtained on the model evaluation.

5. Model Deployment

The model is deployed using the flask framework. Firstly, the model is saved into a pickle file named “RF.pkl”. An API in the flask environment is created which enables the model to allow user input and output. The saved pickle file is loaded onto the flask app created. An app route is created which is rendered to the front-end as a home page and therefore the app is ready for user interaction. Addition to the deployment of the model for the prediction of depression, for further evaluation of depression, a well curated questionnaire is created with four options, which users can choose from and after submitting the questionnaire, a depression score is calculated. Depending upon the score obtained, the user is intimated what level of depression he/she might be going through. The depression levels calculated are of early stage, mild , moderate and severe. The score is shown alongside remedies and a form is provided for the user to give in their contact details which can be further used to connect them to suitable psychiatrists for therapy sessions.

B) Algorithm:

The algorithm used is “Random Forest Classifier”, which is a"supervised machine learning” algorithm .To obtain an optimal output, several methodologies are used, one of them used is methodology of bagging, where we combine different types of learning models to obtain an optimal result. Using the above-mentioned approach, a group of 'decision trees' forms out a “forest”, which adds a lot of unpredictability to the model. When a node is splatted, it serves as a preeminent feature from a random group of characteristics. As a result, a lot of variation is observed, which eventually leads to an optimized model. Because of the default hyperparameters which it employs to frequently yield accurate predictions, it is a useful algorithm. It decreases overfitting and the amount of time consumed training the model. A sample is randomly selected from the dataset and then trees are created individually for each and every sample to obtain result from each tree and then voting is performed for each and every predicted result and eventually the highest voted result is selected as the final result.

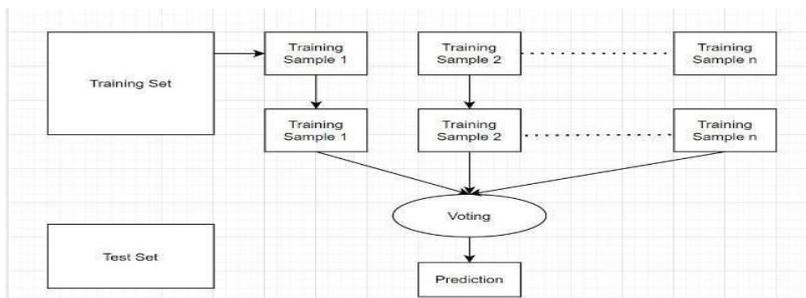


Figure 2: Working of Algorithm



IV. EXPERIMENTATION AND RESULTS

1. The correlation matrix is more like a collection of correlation coefficients for distinct variables. It demonstrates the correlation between all feasible value pairings in a table. Every cell in the table displays the correlation coefficient, and features sharing one as correlation coefficient are highly correlated.

2.

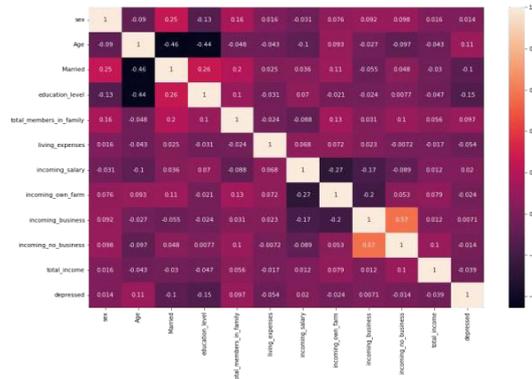


Figure 3: Correlation Matrix for our model

3. Confusion Matrix: The confusion matrix is a kind of “table layout” which allows a statistical visualization of the performance of the algorithm. It is divided into four different combinations spread out with respect to the model’s predicted and actual values. The combinations are as follows:-

If the model predicts the person is depressed and the person is depressed, it is true positive

If the model predicts the person is depressed, but the person is not depressed, it is false positive

If the model predicts the person is not depressed and the person is not depressed, it is true negative
The model predicts the person is not depressed, but the person is depressed., it is false negative

- Accuracy- “The proportion of overall predictions that are correct.”
- Precision- “The ratio of correctly predicted positive observations to overall predicted positive observations is known as precision.”
- Recall- “Defined as the measure, how one correctly identifies true positives in our model. It is also known by the name of sensitivity.”

Specificity: defined as the factor which measures a model's ability to predict true negatives in each available category is called specificity.

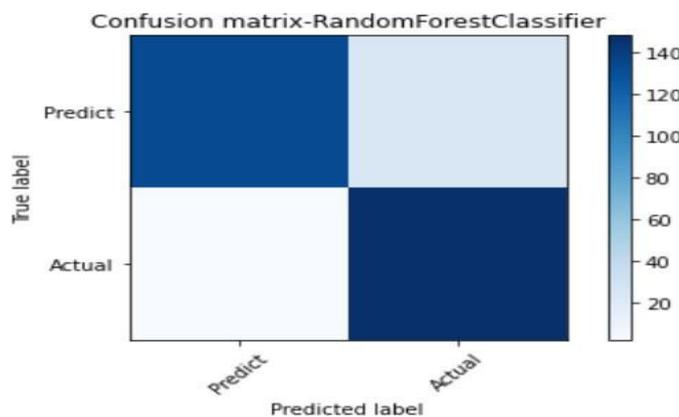


Figure 4: Output for our confusion matrix



V. RESULTS

Depression prediction model entails a thorough examination of the dataset and an understanding of the relationship between factors through the use of Exploratory Data Analysis. We discovered that the Random Forest Classifier model performed the best based on our performance matrix, which included accuracy of 93.28%, precision of 92%, sensitivity of 84%, F1 score of 92%, and recall of 92%. To show the easy flow of our prediction model, a Web application is created to facilitate user-model interaction. When the user is worried and unable to determine the core cause of mental health problems and seek medical care, the model will be invaluable since patients may be simply and properly classified as depressed or not using socio-demographic characteristics to identify mental health disorders.

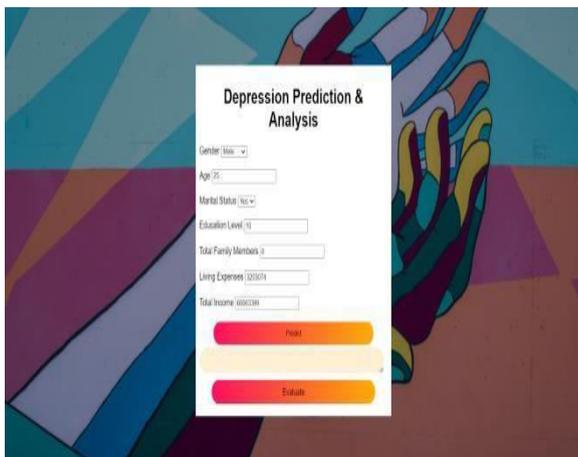


Figure 5: Depression Predictor

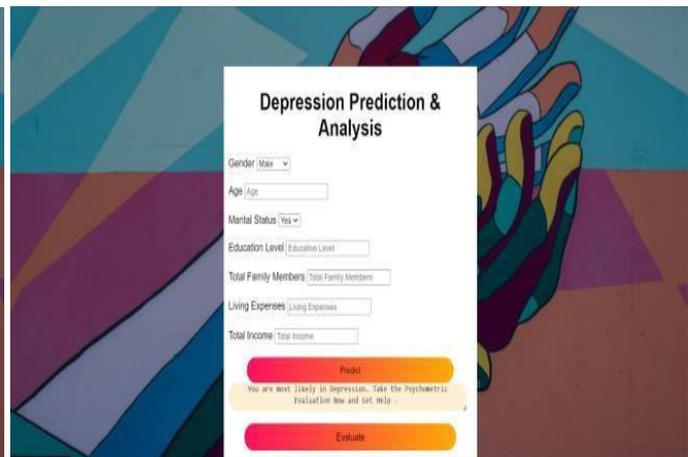


Figure 6: Prediction Result



Figure 7: Psychometric Evaluation

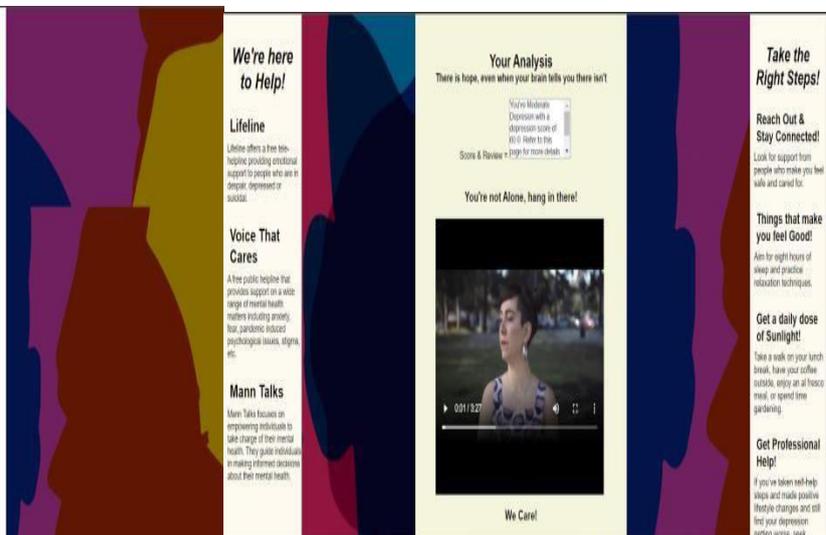


Figure 8: Remedies

VI. CONCLUSION AND FUTURE WORK

The focus of the project is to demonstrate how marital status, gender, age, educational level family members, and income level are among socio-demographic aspects to consider depression. These characteristics had an impact on the health of several people. The conclusions we have derived from this approach are that it's crucial to concentrate on a person's socio-cultural background in terms of improving forecast accuracy and aid therapy sessions. This substantial link between depression and socioeconomic position will be crucial in diagnosing and addressing the condition. This study is anticipated to aid not only in the management of MDD, but also in the more precise identification of patients based on their socio-demographic characteristics



REFERENCES

1. M. E. Aragon, A. P. Lopez-Monroy, L. -C. G. Gonzalez-Gurrola and M. Montes, "Detecting Mental Disorders in Social Media Through Emotional Patterns - The case of Anorexia and Depression", in *IEEE Transactions on Affective Computing*, doi: 10.1109/TAFFC.2021.3075638.
2. Ms. Sumathi M.R., Dr. B. Poorna, "Prediction of Mental Health Problems Among Children Using Machine Learning Techniques", *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 1, 201.
3. Alsagri, Hatoon & Ykhlef, Mourad (2020), "Machine Learning-based Approach for Depression Detection in Twitter Using Content and Activity Features".
4. Sharma, Meera & Mahapatra, Sonok & Shankar, Adeethyia & Wang, Xiaodi., "Predicting the Utilization of Mental Health Treatment with Various Machine Learning Algorithms". *WSEAS TRANSACTIONS ON COMPUTERS*. 19. 285-291. 10.37394/23205.2020.19.34
5. Islam, Md Rafiqul et al. "Depression detection from social network data using machine learning techniques." *Health information science and systems* vol. 6,1 8. 27 Aug. 2018,
6. Bonful, H.A., Anum, A, "Sociodemographic correlates of depressive symptoms: a cross-sectional analytic study among healthy urban Ghanaian women". *BMC Public Health* 19, 50 (2019).
7. Y. E. Alharahsheh and M. A. Abdullah, "Predicting Individuals Mental Health Status in Kenya using Machine Learning Methods", 2021 12th International Conference on Information and Communication Systems (ICICS), 2021, pp. 94-98, doi: 10.1109/ICICS52457.2021.9464608.
8. W. C. de Melo, E. Granger and A. Hadid, "Depression Detection Based on Deep Distribution Learning", 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 4544-4548.
9. M. De Choudhury, S. Counts, and E. Horvitz, "Social media as a measurement tool of depression in populations.," In *Proceedings of the 5th Annual ACM Web Science Conference*, 2013.
10. S. Burdisso, M. Errecalde, and M. Montes-y Gómez, "A text classification framework for simple and effective early depression detection over social media streams," *Expert Systems With Applications*, Vol. 133, 2019.
11. G. Coopersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in twitter," *Workshop on Computational Linguistics and Clinical Psychology*, 2014.
12. S.Kannadhasan and R.Nagarajan, Development of an H-Shaped Antenna with FR4 for 1-10GHz Wireless Communications, *Textile Research Journal*, DOI: 10.1177/00405175211003167, March 21, 2021, Volume 91, Issue 15-16, August 2021
13. B. Saha, T. Nguyen, D. Phung and S. Venkatesh, "A Framework for Classifying Online Mental Health-Related Communities With an Interest in Depression," in *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1008-1015, July 2016
14. M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, 2017, pp. 858-862202



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details