JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Depression Detection Using Machine Learning

Omkar Singh[1], Femenca Noronha[2] , Esha Makwana[3], Mehul Kahar[4]

[1]H.O.D (DS), Department of DS, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

[2]Assistant Professor of DS , [3,4]PG Student, Department of DS, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

omkarsingh@tcsc[1], nfemenca@gmail.com[2], eshamakwana999@gmail.com[3],

kaharmehul@yahoo.com[4]

Abstract : The Depression is the major issues in these World, million of people are suffering from Mental issues due to unavailability of treatment and services of depression detection. Depression is the most common disease now days Depression is a significant mental health disorder that affects millions of people worldwide. Detecting depression at an early stage is critical for timely intervention and treatment.The Main objective of these research to recognized the Symptoms by creating the application, by using the several algorithm like Random Forest, Naive Bayes,and Support Vector Machine and with accuracy of 85%, 72.22%, 80%.The data set is in CVS files.

***IndexTerms* - Depression Detection, Random Forest, Naive Bayes, Support Vector Machine, Mental issues.**

Introduction: The Depression is the major issues in these World, million of people are suffering from Mental issues due to unavailability of treatment and services of depression detection. Depression is the most common disease now days Depression is a significant mental health disorder that affects millions of people worldwide. Detecting depression at an early stage is critical for timely intervention and treatment. Traditional methods of diagnosing depression rely on clinical assessments, interviews, and questionnaires, which can be subjective and time-consuming. With the growing prevalence of mental health issues, there is a need for more objective, efficient, and salable methods for detecting depression. It occur because of Mental issues, family problems, medical issues , stress in daily life these are the main issues. Depression is different mood and loss the interest in every thing and tried from your Life. Machine Learning could change the life easy to recognized the Symptoms by creating the application, by using the several algorithm like Random Forest, Naive Bayes and Support Vector Machine . Depression is a massive mental health disorder that affects millions of people worldwide, leading to significant challenges in personal, social, and occupational functioning. Traditional methods of diagnosing depression often rely on self-reported symptoms and clinical interviews, which can be subjective and time-consuming. With the advent of digital health technologies and the increasing availability of data, there has been a growing interest in leveraging machine learning (ML) techniques to enhance the detection and diagnosis of depression. Machine learning offers the potential to analyze vast amounts of data, identifying patterns and correlations that may not be apparent through traditional diagnostic methods. By processing data from various sources such as social media activity, speech patterns, wearable sensors, and clinical records, ML models can assist in identifying individuals at risk of depression, often before clinical symptoms become severe. The application of machine learning in depression detection typically involves several key steps: data collection, feature extraction, model training, and evaluation. Data collection may include text data from online forums, voice recordings, or physiological data from wearable devices. Feature extraction involves identifying relevant indicators of depression, such as linguistic cues, changes in sleep patterns, or variations in social behavior. These features are then used to train ML models, such as Support Vector Machines (SVM), Random Forests, and Naive Bayes which can classify or predict the likelihood of depression. The integration of machine learning in mental health care offers numerous advantages, including early detection, continuous monitoring, and personalized treatment recommendations. However, it also presents challenges, such as ensuring data privacy, addressing ethical concerns, and managing the accuracy and interpret ability of the models. In this research, we explore the current methodologies and technologies used in machine learning for depression detection. We aim to evaluate the effectiveness of various ML algorithms, understand the types of data most predictive of depression, and discuss the potential implications for future mental health care. By advancing the application of machine learning in this domain, we hope to contribute to more timely and accurate identification of depression, ultimately improving outcomes for those affected by this condition.

Literature Review: M.Deshpande and V. Rao conducted the research on "Depression detection using emotion artificial intelligence," in the year 2017 and level of International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India they used the Support Vector Machine and Naive Bayes algorithm with has 79% and 83% accuracy.[1]. W. C. de Melo, E. Granger and A. Hadid, has conducted the research on "Depression Detection Based on Deep Distribution Learning,"in the year 2019 has published in IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan has used on the deep learning loss obtained[2]. A. Teles et al. Has conducted the research on the "Mobile Mental Health " A Review of Applications for Depression Assistance," int the year 2019 in IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Cordoba, Spain, has worked on the review of the application for depression analysis has a worek on it they made the analysis by using Histogram, Plot and the diagram for the analysis the depression Assistance has made the data in the csv file.[3]. V. Jain, D. Chandel, P. Garg and D. K. Vishwakarma, has conducted the research on the "Depression and Impaired Mental Health Analysis from Social Media Platforms using Predictive Modelling Techniques," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytic and Cloud) (I-SMAC), Palladam, India which has the analysis and precdictive based on the data has also make the anlaysis on the data and using the histogram , plot and the model data created.[4]. K. A. Govindasamy and N. Palanichamy, has conducted the research on the "Depression Detection Using Machine Learning Techniques on Twitter Data," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021 has work on the Twitter Data by using the Machine Learning has used the Naive Bayes and NBTree has work on the 1000 data and the accuracy has 92.34 and also has the graph to show the analysis and prediction.[5]

Methodology & Experimentation: Our first move is to select the topic, we had collect the data by feeling the form we had collect the primary datasets from family, Friends.We started working on the main problem in depression mainly like to know the basic and neutral problem.We clean the data and worked into the training datasets. Then we work into the data creating the models the algorithm we had work on it is Naive Bayes, Random Forest and Support Vector Machine. The data is in CSV format.Random Forest algorithm is a powerful tree learning technique in Machine Learning. It works by creating a number of Decision Trees during the training phase. Each tree is constructed using a random subset of the data set to measure a random subset of features in each partition. This randomness introduces variability among individual trees, reducing the risk of overfitting and improving overall prediction performance.In prediction, the algorithm aggregates the results of all trees, either by voting (for classification tasks) or by averaging (for regression tasks) This collaborative decision-making process, supported by multiple trees with their insights, provides an example stable and precise results. Random forests are widely used for classification and regression functions, which are known for their ability to handle complex data, reduce over fitting, and provide reliable forecasts in different environments.Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other. To start with, let us consider a datasets.

One of the most simple and effective classification algorithms, the Naïve Bayes classifier aids in the rapid development of machine learning models with rapid prediction capabilities. Naive Bayes algorithm is used for classification problems. It is highly used in text classification. In text classification tasks, data contains high dimension (as each word represent one feature in the data). It is used in spam filtering, sentiment detection, rating classification etc. The advantage of using naive Bayes is its speed. It is fast and making prediction is easy with high dimension of data. This model predicts the probability of an instance belongs to a class with a given set of feature value. It is a probabilistic classifier. It is because it assumes that one feature in the model is independent of existence of another feature. In other words, each feature contributes to the predictions with no relation between each other. In real world, this condition satisfies rarely. It uses Bayes theorem in the algorithm for training and prediction has shown in the given figure.

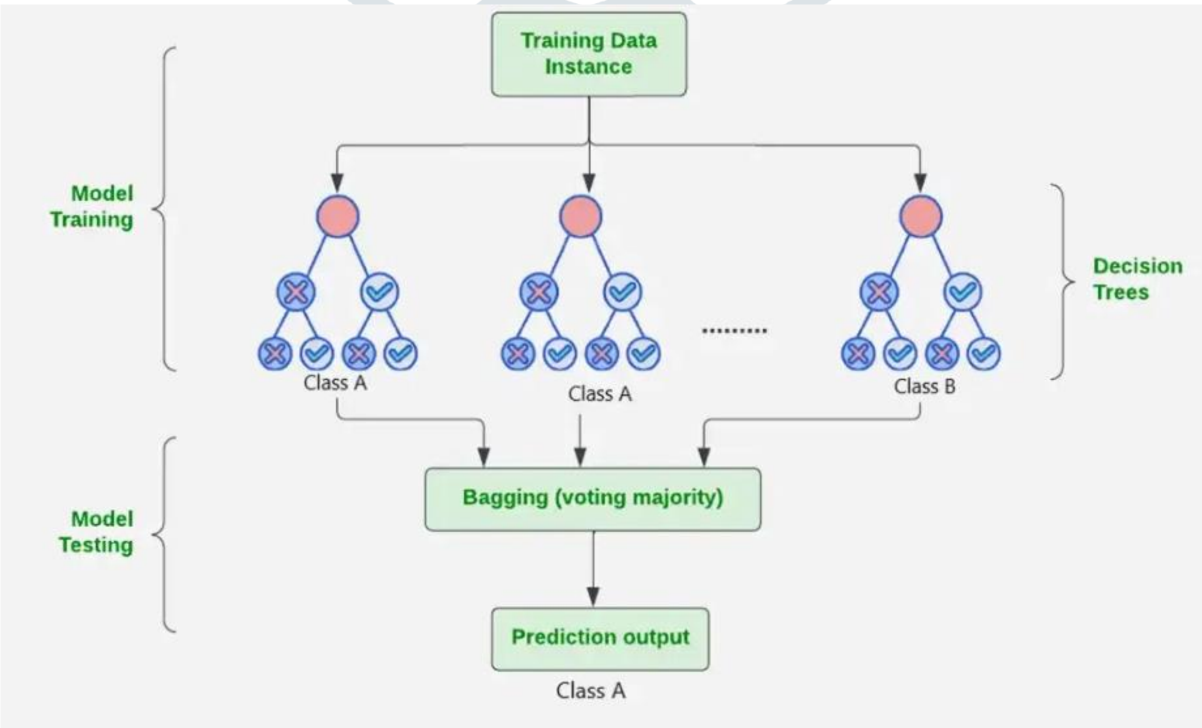

Fig.1 Flow of the Random Forest

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Name	Age	Marital Stat	Education L	Number of	Smoking Sta	Employmen	Alcohol Con	Dietary Hab	Sleep Patter	History of M	History of S	Family Histc	Medical Co
2	Krita shah	26	Married	High School	0	Current	Employed	High	Healthy	Poor	No	Yes	No	No
3	mehul yada	40	Divorced	Bachelor's D	3	Former	Employed	High	Healthy	Good	No	Yes	Yes	No
4	Arun Mehta	18	Single	High School	0	Non-smoke	Employed	Moderate	Unhealthy	Good	No	No	No	No
5	kriti Patel	44	Married	High School	2	Non-smoke	Unemploye	Low	Moderate	Fair	No	Yes	Yes	Yes
6	Raghav Basu	48	Married	Bachelor's D	0	Former	Unemploye	Moderate	Healthy	Good	No	Yes	No	No
7	Payal Das	70	Married	Associate D	3	Former	Employed	High	Unhealthy	Poor	No	No	No	No
8	mahendra B	69	Married	Bachelor's D	1	Non-smoke	Employed	Low	Unhealthy	Fair	No	Yes	No	Yes
9	Akash yadav	42	Married	High School	2	Current	Employed	Low	Moderate	Good	No	No	No	No
10	Ritik Chopra	22	Single	High School	0	Non-smoke	Employed	Low	Unhealthy	Fair	No	Yes	No	No
11	Avik Shah	43	Widowed	Master's De	0	Non-smoke	Employed	Moderate	Moderate	Good	No	No	No	No
12	Priyal Bhatia	60	Married	Bachelor's D	0	Non-smoke	Employed	Moderate	Healthy	Good	Yes	Yes	No	No
13	Alan Garcia	31	Married	PhD	1	Current	Employed	High	Unhealthy	Poor	No	No	No	No
14	Alan Jones	68	Married	Associate D	1	Former	Employed	Moderate	Moderate	Poor	No	No	Yes	Yes
15	Alex Smith	47	Married	Associate D	2	Former	Employed	Moderate	Unhealthy	Fair	Yes	No	No	No
16	Raj Mehta	77	Married	Bachelor's D	2	Non-smoke	Employed	Low	Unhealthy	Poor	Yes	No	No	No
17	Rishi shah	45	Divorced	Bachelor's D	0	Current	Unemploye	Low	Unhealthy	Fair	No	No	No	Yes
18	Ritik Mehta	69	Widowed	Master's De	0	Non-smoke	Employed	Moderate	Unhealthy	Good	No	Yes	No	No
19	Usha Gupta	41	Married	Bachelor's D	3	Non-smoke	Unemploye	Low	Unhealthy	Fair	Yes	No	No	No
20	Suraj Yadav	74	Married	Bachelor's D	0	Former	Employed	High	Unhealthy	Fair	Yes	No	No	No

Fig.2 Datasets

Result: In this research we had used the Machine Learning such as Random Forest, Naive Bayes and Support Vector Machine with these has 85%, 72.22%, 80% these accuracy.These research show the most important in the Society of solving these issues we had also used the three model. These help to improving the mental health and has the Naive Bayes and Support Vector Machine with these has 85%, 72.22%, 80% these accuracy shown in the fig.3, Fig4, fig 5.

```
<ipython-input-1-9fc659ac2a37>:8: FutureWarning: DataFrame.fillna with 'method' is deprecated and will rai
  data.fillna(method='ffill', inplace=True)
Accuracy: 85.00%
Confusion Matrix:
 [[15  2]
 [ 1  2]]
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.88      0.91        17
           1       0.50      0.67      0.57         3

    accuracy                           0.85        20
   macro avg       0.72      0.77      0.74        20
weighted avg       0.87      0.85      0.86        20

Accuracy: 85.00%
```

Fig.3.Accuracy of Random Forest

```
Accuracy: 72.22%
Confusion Matrix:
 [[64  2]
 [23  1]]
Classification Report:
              precision    recall  f1-score   support

           0       0.74      0.97      0.84        66
           1       0.33      0.04      0.07        24

    accuracy                           0.72        90
   macro avg       0.53      0.51      0.46        90
weighted avg       0.63      0.72      0.63        90

Accuracy: 72.22%
<ipython-input-2-6f1b599cdc6c>:8: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  data.fillna(method='ffill', inplace=True)
```

Fig.4.Accuracy of Naive Bayes

```
Accuracy: 80.00%
Classification Report:
              precision    recall  f1-score   support

          No       0.80      1.00      0.89         8
         Yes       0.00      0.00      0.00         2

    accuracy                           0.80        10
   macro avg       0.40      0.50      0.44        10
weighted avg       0.64      0.80      0.71        10

Confusion Matrix:
[[8 0]
 [2 0]]
<ipython-input-3-077341b8f896>:25: FutureWarning: DataFrame.fillna with 'method' is deprecated and wil
  X = pd.DataFrame(X).fillna(method='ffill').values #Fill missing values after Label Encoding
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1471: UndefinedMetricWarnin
  _warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1471: UndefinedMetricWarnin
  _warn_prf(average, modifier, msg_start, len(result))
```

Fig.5.Accuracy of Support Vector Machine

Conclusion: Finally, the purpose of these research depression detection emphasizes that machine learning offers a powerful tool to identify depression more accurately and earlier than traditional methods. By analyzing various forms of data like text, voice, and physiological signals, these models can uncover patterns that may indicate depression. Despite its benefits, the approach comes with challenges, such as ensuring data privacy, avoiding biases, and making sure the models are fair and reliable for all users. Ultimately, machine learning is a supportive tool that can enhance clinical diagnosis and treatment, but human and ethical considerations remain crucial for its successful application. The depression detection using machine learning highlights the transformation potential in mental health diagnostics. Machine learning models can analyze vast amounts of data—such as text, speech, and physiological indicators—to identify patterns indicative of depression with higher accuracy than traditional methods. These models can detect subtle signs of depression early, providing an opportunity for timely intervention and personalized treatment plans.However, despite the benefits, there are challenges that need to be addressed. Data privacy is a key concern, as sensitive personal information is used to train the models. Ensuring the fairness of the algorithms is also essential, as biases in the data can lead to inaccurate predictions, particularly for underrepresented groups. Moreover, while machine learning tools are highly effective, they should be seen as complementary to, rather than a replacement for, human clinicians.Ethical considerations around data usage, model transparency, and accountability must be prioritized to ensure the responsible deployment of these technologies. In conclusion, machine learning has the potential to revolutionize depression detection, but its success depends on balancing technological advances with ethical practices, ensuring that these tools are both effective and equitable for all individuals seeking mental health care.

Reference:

1) M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 2017, pp. 858-862, doi: 10.1109/ISS1.2017.8389299.

2) W. C. de Melo, E. Granger and A. Hadid, "Depression Detection Based on Deep Distribution Learning," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 4544-4548, doi: 10.1109/ICIP.2019.8803467.

3) A. Teles et al., "Mobile Mental Health: A Review of Applications for Depression Assistance," 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Cordoba, Spain, 2019, pp. 708-713, doi: 10.1109/CBMS.2019.00143.

4) V. Jain, D. Chandel, P. Garg and D. K. Vishwakarma, "Depression and Impaired Mental Health Analysis from Social Media Platforms using Predictive Modelling Techniques," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2020, pp. 855-860, doi: 10.1109/I-SMAC49090.2020.9243334.

5) K. A. Govindasamy and N. Palanichamy, "Depression Detection Using Machine Learning Techniques on Twitter Data," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021, pp. 960-966, doi: 10.1109/ICICCS51141.2021.9432203.