

A Model for Predicting Human Depression using Machine Learning Algorithm

J. Shanthalakshmi Revathy, N. Uma Maheswari, S. Sasikala

Abstract---In recent days people share their images in many social media and also they are not certain about the privacy. This leads to many issues. In light of these incidents, a tool is needed for providing access control method during user's file sharing. To do that, Adaptive Privacy Policy Prediction (A3P) system is proposed for setting the privacy parameters to their images. The image data, user's social context and their meta data are used in the privacy setting preferences. The author proposed a two-stage framework for determining and setting the existing privacy policy in accordance to the user's history over the website. The proposed solution depends on the image categories and the social features of the user. Also, it depends on the user's privacy data and the relevant features. From the experimental and simulation results it is proved that the privacy preservation is one of the important task in internet applications.

Keywords---Adaptive Privacy Policy Prediction, privacy preference

I. INTRODUCTION

Major depressive disorder is a disorder affecting individual way of living. Depression has such an impact on people that they start feeling sore of those leisure activities which they used to appreciate before. Their methodology is negative towards everything and it is a typical issue for each age. The World Health Organization estimates that depression will be the world's second significant illness by 2020. Depression presents with angry outbursts, sleep disturbances, low energy and poor concentration and at its worst, can lead to suicide. But by predicting the depression and taking proper treatment a person can lead a normal, healthy life. Since there is no clinical equipment to detect it and we try to predict depression using machine learning classification algorithms.

II. LITERATURE SURVEY

On referring papers related to the chosen topic many information are gathered. Survey focuses on identifying how various algorithms have been implemented and its performance. The algorithms already implemented are such as Apriori, Bayes Net, Logistics, SMO, Decision table, Multi Layer Perceptron and SVM.

A. "Screening of anxiety and depression among the seafarers using machine learning technology" [1]

This paper was published in 2018 Informatics in Medicine Unlocked by ArkaprabhaSau and IshitaBhakta. Here Catboost, Logistic Regression, Naive Bayes, Random Forest and Support Vector Machines machine learning algorithms were used to predict depression. The survey was based on 470 seafarers. Catboost algorithm performed the best in terms of accuracy (82.6%) and precision (84.1%).

B. "Predicting anxiety and depression in elderly patients using machine learning technology"[2]

This paper was published in 2017 in International Journal of Computer Applications by ArkaprabhaSau and IshitaBhakta. In this paper, five classifiers were compared with respect to four metrics – Accuracy, ROC area, precision and Root Mean Square Error. According to the value of ROC area and RMSE, Bayes Net is better than other four classifiers.

C. "A model for predicting human depression using Apriori algorithm" [3]

This paper was published in 2014 International Conference on Information Technology by Lambodar Jena and NarendraK.Kamila. They have used Apriori algorithm and this model predicts the presence of depression at work environment. They have mentioned this work can be extended for prediction of depression at individual level.

D. "Machine Learning Based Prediction of Depression among Type 2 Diabetic Patients" [4]

Raid M. Khalil and Adel Al-Jumaily published this paper in 2017, 12th International Conference on Intelligent Systems and Knowledge Engineering. Classifier algorithms like Support vector machine (SVM), K-Mean, F-Cmean, and Probabilistic Neural Network (PNN) ML algorithms were compared to predict depression disorder. This paper says Depression is strongly associated with type 2 diabetes, so prediction of depression helps the specialists to detect depression as early as possible. Here, SVM classifier generates more precise results than the others.

E. "Classifying Depression Patients and Normal Subjects using Machine learning techniques" [5]

BehshadHosseinifard, Mohammad Hassan Moradi and Reza Rostami published this paper in 2011 Iranian Conference on Electrical Engineering. Classification algorithms such as SVM and LR are used. Here classification algorithms were used to classify depressive and non depressivepatients.

Revised Manuscript Received on January 2, 2020.

J. Shanthalakshmi Revathy, Assistant Professor, Department of Computer Science and Engineering, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India

N.Uma Maheswari, Professor, Department of Computer Science and Engineering, PSNA college of Engineering and Technology, Dindigal, Tamil Nadu, India

S. Sasikala, Assistant Professor, Department of Computer Science and Engineering, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India

Highest prediction accuracy of 88.6% was obtained with SVM classifier.

F. “A Study and Comparison of prediction algorithms for Depression Detection among millennials-A Machine Learning Approach” [6]

This paper was published in 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication. Machine learning classifiers were used here. This paper focuses on the study of three major methods used for prediction of depression in human - using Machine Learning classifiers and WEKA using imaging, ML method and risk factors.

III. DATASET

The dataset is collected from Open Sourcing Mental Illness (OSMI). A survey based on individuals in work place is the dataset.

This dataset contains the following data: [25 Features]

- Age
- Gender
- Country
- state: Which state or territory do you live in if you live in the United States?
- self_employed: Are you an independent worker?
- Familyhistory: Do you have a mental illness family history?
- Treatment: Did you seek mental health treatment?
- work_interfere: Do you feel it interferes with your work if you have a mental health condition?
- no_employees: How many employees does your organization or company have?
- remote_work: Do you work atleast 50 percent of the time remotely (outside an office)?
- tech_company: Is your employer a tech company/organization primarily?
- benefits: Does your employer provide benefits for mental health?
- care_options: Do you know the mental health care options that your employer offers?
- wellness_program: as your employer ever discussed mental health as part of a wellness program for your employees?
- seek_help: Does your employer provide resources to learn more about issues of mental health and how to seek assistance?
- anonymity: If you choose to take advantage of their sources of mental health or substance abuse, is your anonymity protected?
- leave: How easy is taking medical leave for a condition of mental health for you?
- mental_health_consequence: Do you think it would have negative effects to discuss same mental health issue with your employer?
- phys_health_consequence: Do you think it would have negative effects to discuss a physical health issue with your employer?
- coworkers: Would you be willing to talk to your colleagues about a mental health issue?
- supervisor: Would you like to talk to your direct supervisor(s) about a mental health issue?

- mental_health_interview: Are you raising a mental health issue in an interview with a potential employer?
- phys_health_interview: Would you like to raise a physical health issue in an interview with a potential employer?
- mental_vs_physical: Do you feel that mental health is taken as seriously as physical health by your employer?
- obs_consequence: Have you heard or noticed negative effects on coworkers in your work place with mental health conditions?

IV. EXISTING SYSTEM

Most of the predictive model based on specific group of working population. Data set taken for training and testing the model is very minimum. But the success of machine learning technology in mental health care screening among them, can be widened to encompass the others in working population. Various algorithms have been used in various models but those models are limited to particular group of people.

V. PROPOSED SYSTEM

Machine learning algorithm such as Support Vector Machine (SVM), Bayes net and catboost are being implemented here to design a model to predict depression.

A. SVM:

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for classification or regression problems. It is mostly used in classification issues, then regression. In this algorithm, every data set is plotted as a point in n - dimensional with the estimation of each element being the estimation of a particular facilitate. Then, grouping can be performed by finding the hyper-plane that separate the two classes great.Support Vectors are essentially the co-ordinates of individual perception. Support Vector Machine is a fringe that isolates the two classes best.

SVM varies from the other grouping calculations by picking the choice limit that expands the separation from all classes ' nearest information focuses. A SVM does not simply discover a limit for a choice; it finds the most ideal limit for a choice

B. Bayes net:

A Bayesian network is a probabilistic realistic model setofrandomvariablesandtheirdependenciesthroughadirected acyclicgraph(DAG). Learning process in Bayesian system is a two phase process: First, take in a system structure from the datasets. Second, take in the likelihood table's from the recipe. Bayes model is anything but difficult to construct and especially valuable for vast informational indexes.

In addition to simplicity it is realized that Naïve Bayes out performs even profoundly advanced characterization strategies. Bayes provides away to calculate $P(c)$ from $P(c)$, $P(x)$ and $P(x)$.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability
Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

P(c) is the posterior probability of class (c, target) given indicator (x, attributes). P(c) is the previously mentioned likelihood of class. P(x) is the likelihood of class given indicator. P(x) is the previously mentioned indicator likelihood

C. Catboost:

CatBoost algorithm is a recently developed open-sourced machine learning algorithm from Yandex.

It can be very well incorporated with deep learning frameworks for example Google’s Tensor Flow and Apple’s Core ML. It can work with an assortment of data types to help and solve a wide scope of issues that most of the businesses are confronting today. To top it up, it gives best-in-class accuracy.

It is especially amazing in two different ways: it conveys state - of - the - art results without broad information preparing ordinarily required by different techniques for AI, and gives incredible out - of - the box support for increasingly spellbinding information arranges that go with numerous business issues.

The library functions admirably with numerous information classifications, including sound, content, picture, and authentic information. Since an angle boosting library depends on this library. Gradient boosting is a ground-breaking AI calculation broadly connected to different sorts of business difficulties, for example, extortion location, proposal things, estimating, and it additionally performs well. In complexity to DL models that need to gain from an enormous measure of information, it can likewise return generally amazing outcomes with moderately less information.

VI. EXPERIMENTAL PROCEDURE

A. Data collection:

Collecting the data sets. It is the initial phase in which the dataset is being collected. The collected dataset is used to train and test the system.

B. Data cleaning:

Removing the missing attributes and considering the common attributes. The collected dataset may contain unknown values, incomplete and inappropriate fields. Those kinds of data have to be removed.

C. Training:

Training the machine with existing data sets using machine learning algorithm. System will be trained using existing dataset. 60% of the dataset is used to train the system.

D. Testing:

Testing the created model using new input data to check the output. The remaining 40% of the dataset will be used to test the system.

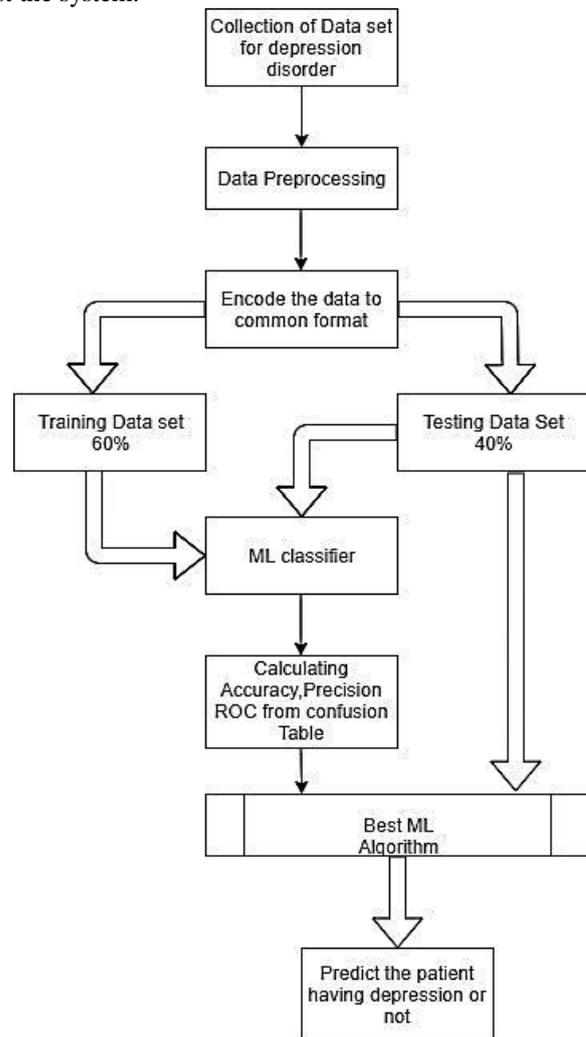


Fig.1: Data Flow Diagram

VII. RESULT

Performance of different machine learning algorithms such as Support Vector Machine (SVM), Bayesian Network and Catboost were compared. ROC Curve, Confusion matrix and classification report are used to check the quality. Metrics such as accuracy, precision, F1- score, error rate, TP, TN, FP, FN are being used for comparison to find which classifier is efficient in terms of accuracy.

$$\begin{aligned} \text{Accuracy} &= ((\text{TP} + \text{TN}) / \text{Total}) * 100 \\ \text{Error_rate} &= ((\text{FP} + \text{FN}) / \text{Total}) * 100 \\ \text{Precision} &= \text{TP} / (\text{TP} + \text{FP}) * 100 \\ \text{Recall} &= \text{TP} / (\text{TP} + \text{FN}) * 100 \\ \text{f1} &= 2 * ((\text{precision} * \text{recall}) / (\text{precision} + \text{recall})) \end{aligned}$$

TP- True positive
TN- True negative
FP- False positive
FN- False negative

A. Confusion Matrix:

A confusion matrix is a table used to describe a classification model's performance (or a "classifier") on a set of test data set for which the true values are already known.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Fig. 2: Confusion Matrix

B. ROC CURVE:

A useful tool when predicting the probability of a binary outcome is the Receiver Operating Characteristic curve or ROC curve. Fig. 3 to 5 shows ROC curve for SVM, BayesNet and Catboost classifier algorithm. Performance of Catboost classifier is best by comparing the ROC curve.

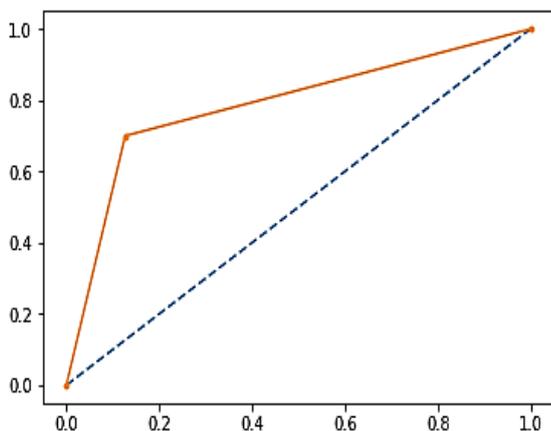


Fig. 3: ROC Curve for SVM

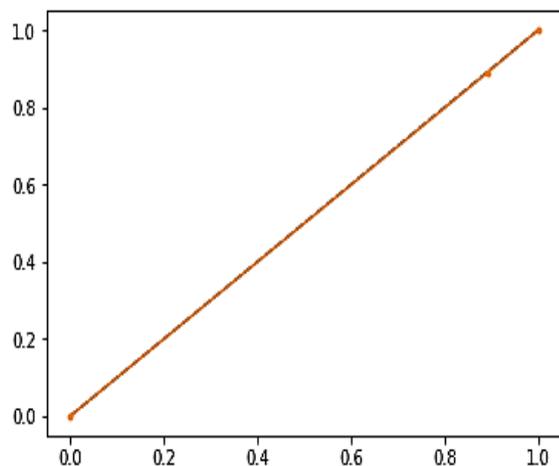


Fig. 4: ROC Curve for BayesNet

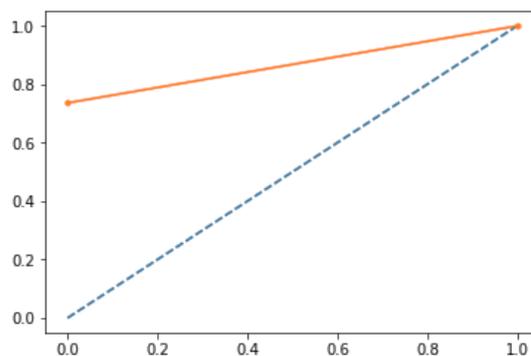


Fig. 5: ROC Curve for Catboost

C. Data visualization of accuracy and error rate using bar chart:

This chart clearly depicts the differences between the 3 different algorithms SVM, BayesNet and Catboost in terms of accuracy and the error rate respectively. From Fig. 6 Catboost shows higher accuracy percentage and less error rate.

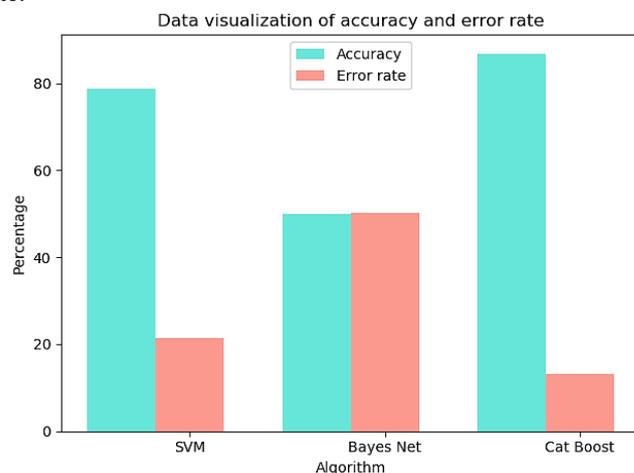


Fig. 6 Comparison of Accuracy and Error Rate

VIII. CONCLUSION

This model is aimed at predicting depression at early stage, with high accuracy. It attempts to predict depressive disorders of an individual. There by can prevent the other diseases rooted from depression at initial stage itself. Here the algorithms used are very stable and are compared to check which classifier produces high accuracy in prediction. This model can be extended to get more accuracy.

REFERENCES

1. ArkaprabhaSau&IshitaBhakta, "Screening Of Anxiety And Depression Among The Seafarers Using Machine Learning Technology", Informatics In Medicine Unlocked , 2018, PP 1-7.
2. ArkaprabhaSau&IshitaBhakta, "Predicting anxiety and depression in elderly patients using machine learning technology", Healthcare Technology Letters, 2017, PP. 238-243.
3. Lambodar Jena &NarendraK.Kamila, "A Model For Predicting Human Depression Using Apriori Algorithm", IEEE/13th International Conference on Information Technology,2014, PP 240-244.

4. Raid M .Khalil & Adel Al-Jumaily, “Machine Learning Based Prediction Of Depression Among Type 2 Diabetic Patients”, IEEE, 2017.
5. Behshadhosseinifard, Mohammad Hassan Moradi& Reza Rostami, “Classifying Depression Patients And Normal Subjects Using Machine Learning Techniques”, IEEE, 2011.
6. MadhurimaHooda, Aashie Roy Saxena, Dr. Madhulika&BabitaYadav, “A Study And Comparison Of Prediction Algorithms For Depression Detection Among Millennials-A Machine Learning Approach”, IEEE, 2017, PP 779-783.

AUTHORS PROFILE

J. Shanthalakshmi Revathy, Assistant Professor, Department of Computer Science and Engineering, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India

N. Uma Maheswari, Professor, Department of Computer Science and Engineering, PSNA college of Engineering and Technology, Dindigal, Tamil Nadu, India

S. Sasikala, Assistant Professor, Department of Computer Science and Engineering, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India