



Literature Survey on Comparison of Machine Learning Algorithms Used to Predict Students' Engagement on Social Media Platforms During Covid-19

Mr. R A Jamadar¹, GauriPatil²

1. Professor, Department of Information Technology, AISSMS's Institute of Information Technology, Pune-411001, INDIA

2. TE.(Information Technology), AISSMS's Institute of Information Technology, Pune-411001, INDIA

ABSTRACT

Covid-19 was undoubtedly an atypical situation for all & specially for students to sustain productively. Students engaged with online classes, tutorials, practical's, lab, other study related projects along with direct and indirect engagement in social media. So, predicting students' social engagement during the pandemic becomes more challenging due to the large volume of data in databases of numerous social media platforms. A systematical literature review on predicting student engagement on social media platforms by using machine learning and deep learning algorithms is proposed. The main objective of this paper is to provide an overview on the algorithms and compare their performances that have been used to predict students' engagement.

Keywords: Machine Learning, Deep Learning, Classification

1. Introduction to Data Mining

Foreestablishment of problem-solving approach through understanding data cleaning, clustering, and classification, data mining is used. The technology is pervasive in nature to summarize and analysis of data from different directions and dimensions. For extracting features from raw data, Machine Learning is very popular. It is also known for being structurally hierarchical and derive prediction. It supports to manage heterogeneous information from the real world with structuring dataset.

1.1 Motivation

- To compare students' engagement on most used social networking sites by them as per a survey conduction.
- Platforms like LinkedIn, Facebook, Instagram, Quora, Reddit, Snapchat, Telegram, Twitter, WhatsApp, YouTube & etc. as per student survey, are taken into consideration.

1.2 Aim and Objective

Aim

This paper aims compare performance of few machine learning algorithms in the reference of student social engagement during the Covid-19 pandemic.

Objectives

- Data acquisition, feature abstraction, feature alignment, noise removal, splitting and labeling data to utilize further.

- The numerical and categorical data classified properly before using to next level.
- The collected data is preprocessed well for applying various Machine Learning algorithms to derive better accuracy.

2. Literature Review

To support this work, several related studies have been analyzed.

1. Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A case study of small-scale agricultural load.

This paper has presented the use of Auto Regressive Integrated Moving Average (ARIMA) method for forecasting of seasonal time series data. Comparing different models and selecting the optimum ARIMA model by analyzing their respective Akaike and Bayesian information.

Limitation: Method can be applied only when the time series data is stationary.[1]

2. Deep learning models for plant disease detection and diagnosis.

Specialized deep learning models were developed, based on specific convolutional neural networks architectures, for the identification of plant diseases. The training of the models was performed using an openly available database of 87,848 photographs.

Limitation: The model identified correctly the plant species but did not accurately detect the existing plant disease.[2]

3. Analysis of Mobile App Coverage Popularity with ML Algorithms.

Applying ML and DL on sensed data the form the various app released from 2020 as app categorization & rating, other prediction like instalment of app as per category and app detailing can be made

3. Algorithm

Naïve Bayes

- Naïve Bayes algorithm is known about the precise nature for probability model.
- It shows the predicated social media platform with students' age attribute.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

J48

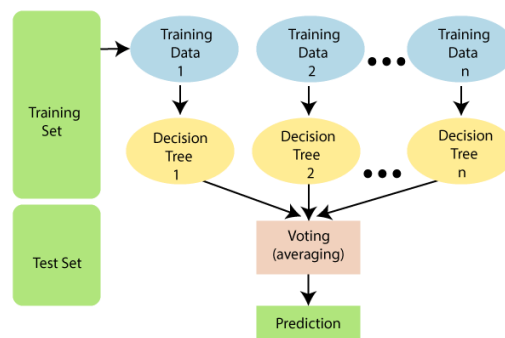
- J48 is one of the most popular in examine categorical data in the concern continuous approach.
- This algorithm act on decision tree basis.

RepTree

- It is one of the fast decision tree learners.
- The regression/decision tree is built up on the structured dataset projected to analyzed.

Random forest

- Random forest is working on the branches of many decisions trees. The individual tree is created with bagging and randomness approach.
- Each tree has its forest to build up the prediction in effective manner rather than any individual tree.



Machine learning Random Forest Algorithm

TOOL USED-

- Weka tool is used to perform various algorithms to be analyzed effectivity.
- To predict the most popular media platform among the survey done in the students with concern of data cleansing, clustering, and classification, performance comparison is approached.

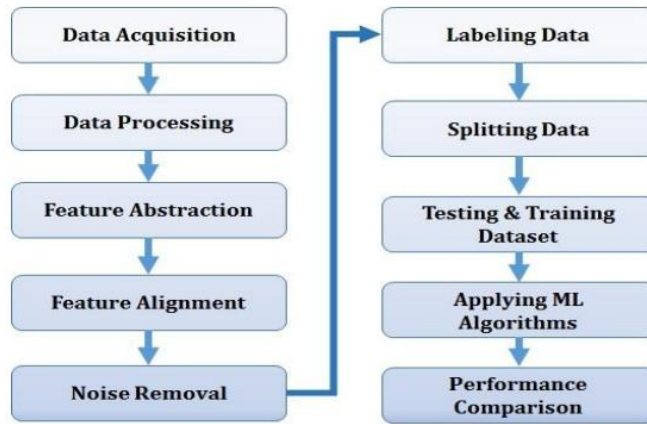
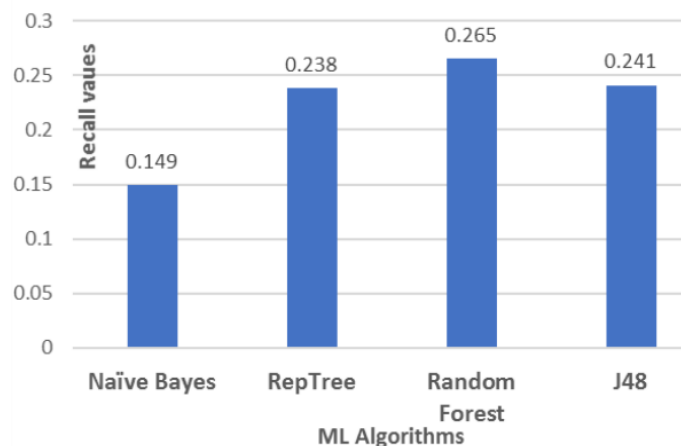


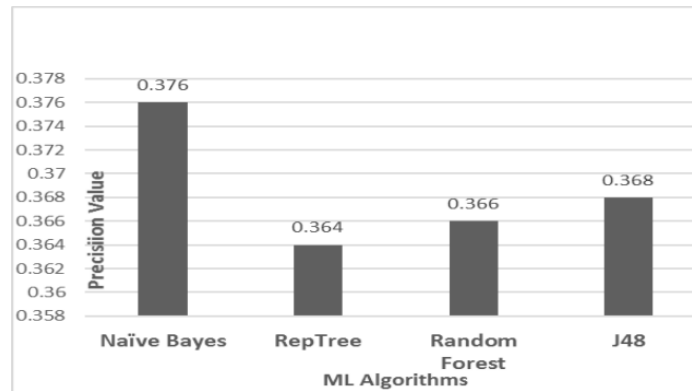
Fig. 1. Flow Diagram

4. Experimental Methodology

1. A dataset of more than 1200+ instances along with more than 30 attributes was collected.
2. The dataset consisted of various attributes including area of residence, age, time spent on online classes, rating of online class experience, medium for online class and time spent on self-study, fitness, sleep, and on social media by each and every student.
3. Decision tree classification algorithms were taken into consideration as well due to their hierarchical nature & diversity along with other algorithms to simulate the prediction of social engagement media during pandemic.
4. The comparison of Naïve Bayes, J48 Tree, RepTree and Random Forest Algorithm is carried on the structured dataset of 1200+ instances.
5. The paper focuses on accuracy, F-measure and time to summarize comparison result.
6. The results of comparison of the four machine learning algorithms with respect to accuracy, F-measure and time were-



Comparison of Recall of Algorithms



Accuracy of various algorithms using precision

TABLE I: COMPARISON OF ACCURACY

<i>Algorithm</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>Time-Taken</i>
Naïve Bayes	0.376	0.149	0.213	0.01
RepTree	0.364	0.238	0.288	0.07
Random Forest	0.366	0.265	0.307	1.3
J48	0.368	0.241	0.291	0.2

5. Conclusion

The applied Machine Learning algorithms helped to generate high recommendations for prediction of students' engagement in various social media platforms.

The study concludes that J48 is the most desirable machine learning algorithm for the same.

REFERENCES

1. S. Atique, V. Roy and S. Bayne, "Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A case study of small scale agricultural load," 2019IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS), Dallas, TX, USA, 2019, pp. 521-524.
2. K. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis", Computers and Electronics in Agriculture, vol.145, February 2018, pp. 311-318.
3. S.Ramesh, D.Vydeki, "Recognition and Classification of Paddy Leaf Diseases using Optimized Deep Neural Network with Jaya Algorithm", Information Processing in Agriculture, KeAi (Chinese roots Global Impact), 2019.
4. G.Webb, J. Boughton "Not So Naive Bayes: Aggregating OneDependence Estimators". Machine Learning, vol. 58, no. 1, 2005, pp.5-24.
5. Kumar, NM Saravana. "Implementation of artificial intelligence in Imparting education and evaluating student performance." Journal of Artificial Intelligence 1, no. 01 (2019): 1 -9
6. Muthukumar, Vignesh, and N. Bhalaji. "MOOCVERSITY-Deep Learning Based Dropout Prediction in MOOCs over Weeks." Journal of Soft Computing Paradigm (JSCP) 2, no. 03 (2020): 140- 152P.Kapoor, Efficient Decision Tree Algorithm Using J48 and Reduced Error Pruning. International Journal of Engineering Research and General Science, vol.3, issue 3,2015, pp. 1613-1621.
7. B. Srinivasan, P.Mekala, "Mining Social Networking Data for Classification Using REPTree", International Journal of Advance Research in Computer Science and Management Studies, vol. 2, issue 10,2014 pp.155-160.
8. H.Trevor; T.Robert; F.Jerome, The Elements of Statistical Learning (2nd ed.). (2008). Springer. ISBN 0-387-95284-5.