Analysis of Mobile App Coverage Popularity with ML Algorithms

Dr. Jigna B Prajapati,
Assistant Prof., Acharya Motibhai Patel Institute of computer studies,
Ganpat University.

Abstract: Current technologies required deep focus from Machine learning to perform better in major heterogeneous industries. Deep learning and Machine learning are becoming popular for accuracy and ease. This paper includes comprehensive analysis of broad area coverage of mobile app with machine learning (ML) and deep learning. This paper also studies the comparison of neural network models. It covers dynamic analysis on usage of various kind of app engaged with different categories with ANN multilayer perceptron, Naïve bayes, J48 tree, REPTree and LMT algorithm. In this paper, we propose and scrutinize major app area and compare their usage coverage capabilities with the various ML approach. Dynamic Analysis is conceded with various attributes of mobile apps which covers the complete data related their commercial aspect. Further Comparison of accuracy, F-measure and time to summarize comparison result. The findings of this paper indicate ML/Deep learning models give high accuracy and outperform for preprocessed mobile app dataset.The finding predicts the most popular category of mobile app using ML.

Keywords: Mobile app, M-app category, Machine Learning, Deep Learning

I. INTRODUCTION

Today's fast growing and super high tech industries face many challenges for their product adaptability. DL & ML play a vital role to help in various production. The commercialization of product need keen focus to be success in market. Every industry come and play in our mobile. They all need direct and indirect app approach to cover the user requirement in different manner. The people are habituated to access mobile app for no reason. As per user interest in using mobile app, the development of mobile app is increased very extensively.

DL is popular for the extraction of features from raw data. Pooling, activation, convolutions, fully connected layers, gates, memory cells and many more are DL features. It is also popular as structural hierarchical and derive prediction. The ML supports to manage heterogeneous information from the real world with structuring dataset [1,2]. We have generated the artificial neural network along with various component of AI to simulate the predication of bread coverage area of used mobile app as per released dated of 2020. The focus analysis of the same paper is to essence on design and deployment of mobile app dataset for forecasting of broad area coverage, most popular area of mobile usage. The collected data is preprocessed for applying various ML and DL algorithms derive better accuracy.

To support this paper, we have also studies the [3]predication of corn yield is covered with deep neural network using environment and genotype data. The sample training dataset training is used for weather& yield prediction. In [4] The paper forecasting the price of agricultural products using Auto ARIMAmodel. It also includes back propagation (BP) network method. Further comparison of accuracy is derived. In [5]The harvest creation in the reference of farmers is discussed with ML. it focused how yield profitability and an investigation can help farmers.In [6] wheat, bajra, rice, pulses the main aim was to achieve increased crop yield for Maharashtra zone. In [7], price forecasting using ANN with FBPfor Mysore district is done. MSE is used to evaluate the accuracy of the model. In [8] the ML is used to detect the plant diseases. In [9] the proposes a deep learning technique for classification of paddy leaf diseases. In [10] the 10-fold cross validation method used to derive higher accurate output then one with the straightforward CNN with 10 or more hidden layers.

II. METHODOLOGY

The row data set is collected and processed to apply various algorithm. The figure:1 describe the architectural model for various phase of actions. The mentioned data mining algorithms are applied on a data set. The data consist of containing different app, category, developers, website, rating, Install, minimum install, maximum install, cost, age rating and some other attributes. The dataset has been made structural to achieve accuracy with help of F-measure. Dynamic analysis done to compare and understand different algorithms. On the basic of various attribute and algorithm, accuracy will be derived with optimal F-measure will be summarizing with the help of Weka tool.

III. IMPLEMENTATION

Artificial neuralnetworks

The Figure:1 describe the visualization of attributes used in the dataset. The dataset includes various attributes including category as Music & Audio, Sports, Finance, Entertainment, Shopping, Educational, Medical, News & Magazines & etc. The dataset comprises of many elements in concern with Category (broad area of app). As shown in fig:2 presents ANN model for multilayer perceptron for the mentioned dataset. It also examines with random noise plotting to analyses what is happening at different spread out of point.

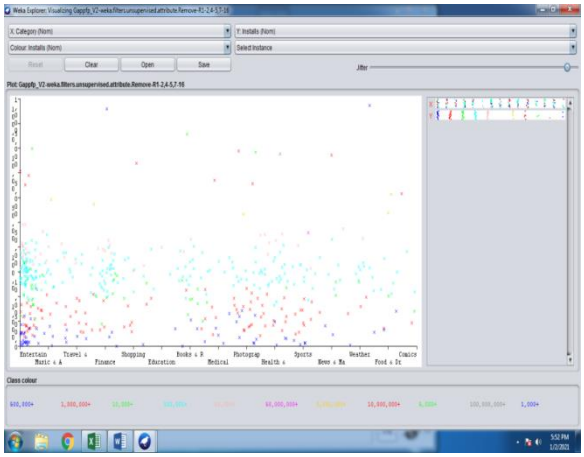

Fig: 1 visualization of attributes

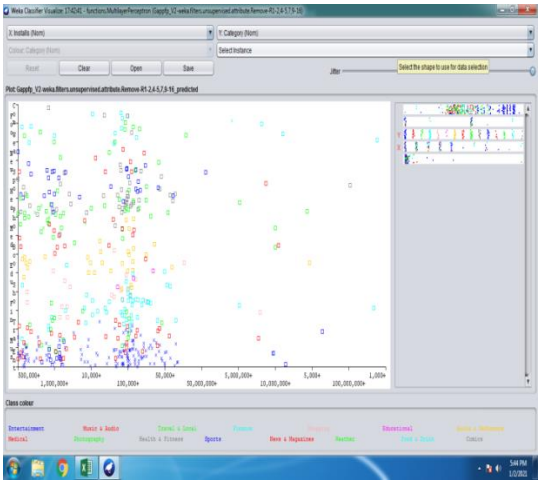

Fig 2: Analysis using ANN multilayer perceptron

NaiveBayes

Figure 3 presents the Naive bayesalgorithm for comparison and reduction. Naïve Bayes algorithm is known about theprecise nature for probability model.

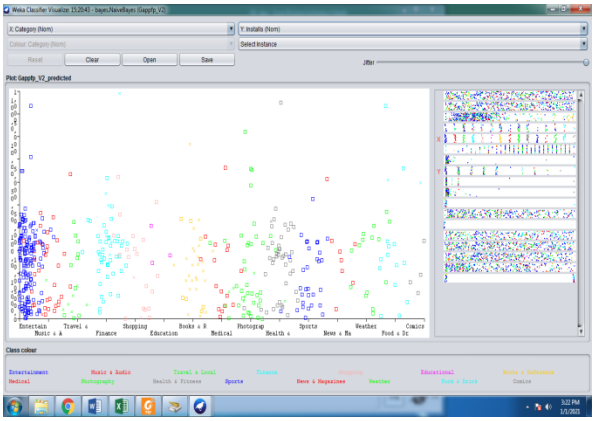

Fig 3: Analysis using Naïve bayes algorithm

J48

Figure:4 represents J48 algorithm.This algorithm act on decision tree basis. It is continuance of Quinlan's ID3 algorithm. An ID3 algorithm can be used to lessening of decision trees, value extents, derivations of rules and so on. J48 algorithm is known for statistical analysis.

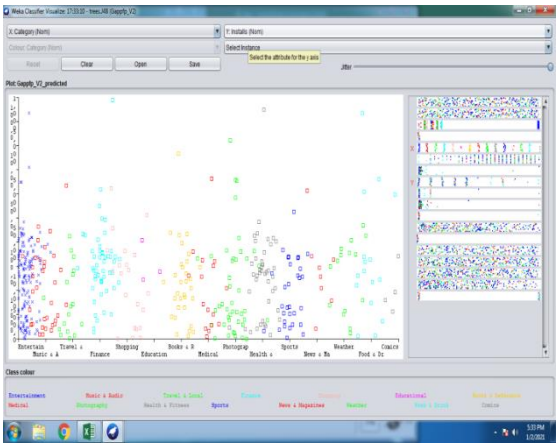

Fig 4: Analysis using J48 tree algorithm

RepTree

Figure 5 present implementation of RepTree algorithm. It is knoan for regression tree for analysis. One of the most highlighting feature of algorithm is fast decision tree.

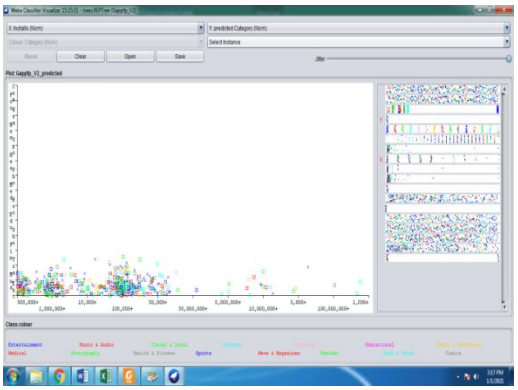

Fig 5: Analysis using REPTree algorithm

LMT

Figure 6 present LMT algorithm. This is supervised algorithm based on classification. The decision tree with Logistic model along with linear regression models.

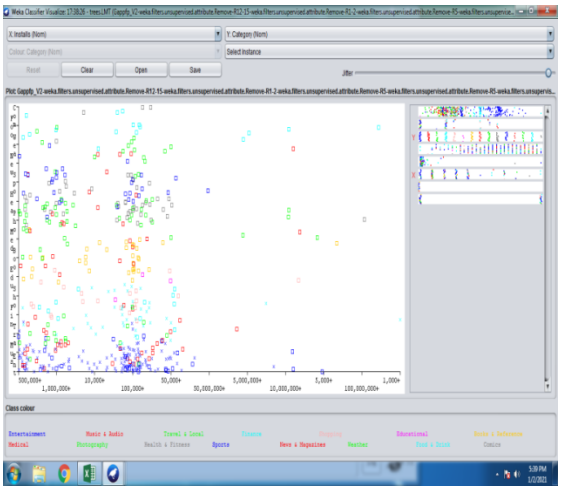

Fig 6: Analysis using LMT algorithm

IV. RESULTS ANDDISCUSSION

The Dataset has been analyzed with cross validation method. The different performance measures are applied to dataset. The table:2 presents the comparative analysis of decision tree classifications & simulation results.

Table 2: Comparison of accuracy

Algorithm	**Precision**	**F1-Measure**	**Recall**	**Time-Taken** (sec)
LMT	0.393	0.521	0.803	0.33
NaïveBayes	0.6	0.072	0.038	0.1
REPTree	0.318	0.483	1.0	0.05
ANN	0.332	0.485	0.902	6.6
J48	0.318	0.483	1.0	0.12

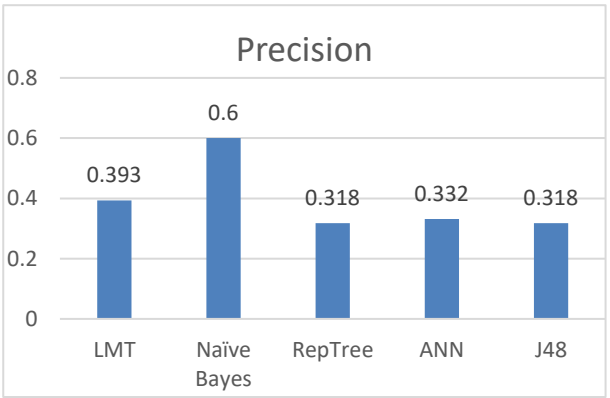

Fig 8: Accuracy of various algorithms using precision

Precision (P): It is to determine of exactness. It is the ration

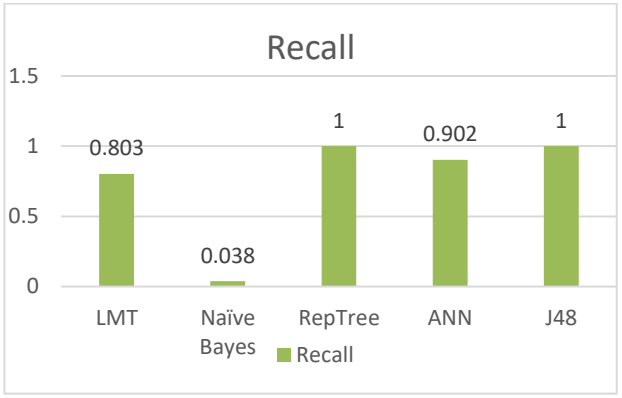

Fig 7: Comparison of accuracy with Recall

Figure 7 represent in terms of RecallReptree& J48 are the better approach in the caparison of LMT .**F-Measure**: The harmonic mean of precision and recall. It is an important measure as it gives equal importance to precision andrecall.

Figure 8 represent the Naïve Bayes is a better algorithm in comparison of rest of others. It has derived the highest and accurate recession. To choose alternate option to it LMT is preferable with concern of Precision. Retiree takes the least score, so it is not very reliable to use for this research.**Recall(R):** Recall is determining of completeness. It is the proportion of positive cases that were correctly recognized to the total number of positive cases. It is also known as sensitivity or true positive rate(TPR).

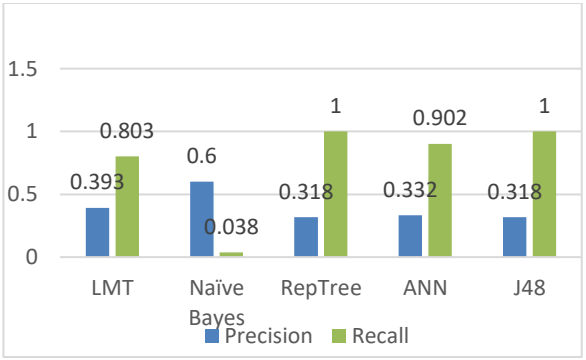

Fig 9: Comparison of accuracy with F-measue

Figure 9 represents the shows the combination of Recall and Precision makes the F-Measure, J48 could be a better algorithm as it has the highest performance in both recall and precision.

Time-Taken

The number of operations executed and specific input are responsible for time factor for any algorithm. As there are long numbers of operations, the longer running time would be resulted. The data set used for more than five hundred instance along with 14 attributes have resulted as below in Fig:10. One can concentrate about the Time taken for the algorithm to process could be one of the criteria to compare the faster algorithm. As per the tanle:2 describe the ANN is slow one and naïve bayes is fast one .

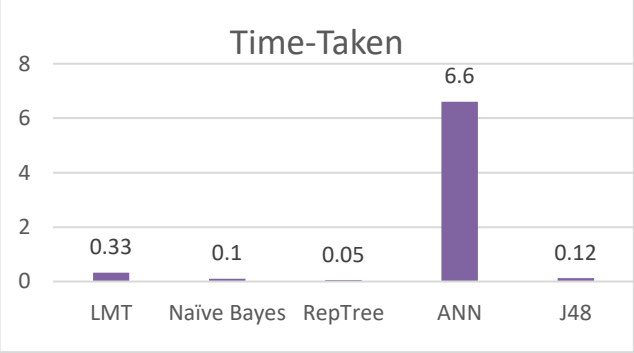

Figure 10: Time taken by algorithms

V. CONCLUSION

We have discussed the economy, practical and easy to develop model in the reference of deep learning and Machine learning. The collected dataset is processed for noise removal. The dynamic analyze is carried out on structural dataset with different parameters and attributes. Applying ML and DL on sensed data the form the various app released from 2020 as app categorization & rating, other prediction like installment of app as per category and app detailing can be made. These ML algorithms is helping to generate high recommendations for correct decision making from the developer to focus the category of app popularity along with its age rating. The above analysis can derive the road are of mobile app is entertainment where users are highly interested to install various types of app.

VI. REFERENCES

[1] AyonDey,"Machine Learning Algorithms: A Review", International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179

[2] M. Bowles, "Machine Learning in Python: Essential Techniques for Predictive Analytics", John Wiley & Sons Inc., ISBN: 978-1-118-96174-2

[3] Khaki, Saeed, and Lizhi Wang. "Crop yield prediction using deep neural networks." In INFORMS InternationalConference on Service Science, pp. 139-147. Springer, Cham, 2019.

[4] Y. Weng, X. Wang, J. Hua, H. Wang, M. Kang and F. Wang, "Forecasting Horticultural Products Price Using ARIMA Model and Neural Network Based on a Large-Scale Data Set Collected by Web Crawler," in IEEETransactions on Computational Social Systems,vol.6,no.3,pp.547-553,June2019.doi:10.1109/TCSS.2019.2914499

[5] Verma, Akanksha, AmanJatain, and Shalini Bajaj. "Crop Yield Prediction of Wheat Using Fuzzy C Means Clustering and Neural Network."International Journal of Applied Engineering *Research* 13, no. 11 (2018):9816-9821.

[6] N. Gandhi, O. Petkar and L. J. Armstrong, "Rice crop yield prediction using artificial neural networks,"*2016IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, Chennai, 2016, pp.105-110.doi: 10.1109/TIAR.2016.7801222

[7] Noureen, 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,"*2019 IEEE 62nd International MidwestSymposium on Circuits and Systems (MWSCAS)*, Dallas, TX, USA, 2019, pp.521-524.doi:10.1109/MWSCAS.2019.8885349

[8] Mohan, Pushpa, and Kiran KumariPatil. "Deep Learning Based Weighted SOM to Forecast Weather and Crop Prediction for Agriculture Application."*International Journal of Intelligent Engineering and Systems* 11, no. 4 (2018):167-176.

[9]] Konstantinos P. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis", Computersand Electronics in Agriculture 145:311-318 · February2018

[10] S. Ramesh , D. Vydeki, "Recognition and Classification of Paddy Leaf Diseases using Optimized Deep Neural Network with Jaya Algorithm", Information Processing in Agriculture,KeAi(Chinee roots Global Impact),