An Empirical Approach for Credit Card Fraud Detection Utilizing Ensemble Classification

Chaitanya Kumar Reddy EC[1], Anjan Babu G[2]

Dept of Computer Science, SV University, Tirupati

***Abstract*—** *The rapid growth in E-Commerce industry has led to an exponential increase in the use of credit cards for online purchases and consequently they have been surged in the fraud related to it. It is vital that credit card companies are able to identify fraudulent credit card transactions so that customers are not charged for items that they did not purchase. In recent years, for banks has become very difficult for detecting the fraud in credit card system. Machine learning plays a vital role for detecting the credit card fraud in the transactions. For predicting these transactions banks make use of various machine learning methodologies, past data has been collected and new features are been used for enhancing the predictive power. The performance of fraud detecting in credit card transactions is greatly affected by the sampling approach on data-set, selection of variables and detection techniques used. This paper investigates the performance of AdaBoost and XGBoost for credit card fraud detection.*

I. INTRODUCTION

Credit card fraud is a huge ranging term for theft and fraud committed using or involving at the time of payment by using this card. The purpose may be to purchase goods without paying, or to transfer unauthorized funds from an account. Credit card fraud is also an add on to identity theft. Fraud in credit card transactions is unauthorized and unwanted usage of an account by someone other than the owner of that account. Necessary prevention measures can be taken to stop this abuse and the behaviour of such fraudulent practices can be studied to minimize it and protect against similar occurrences in the future. In simple terms, Credit Card Fraud is defined as when an individual uses another individuals' credit card for personal reasons while the owner of the card and the card issuer are not aware of the fact that the card is being used. Card fraud begins either with the theft of the physical card or with the important data associated with the account, including the card account number or other information that necessarily be available to a merchant during a permissible transaction.

Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behaviour, which consist of fraud, intrusion, and defaulting. This is a very relevant problem that demands the attention of communities such as machine learning and data science where the solution to this problem can be automated. This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance. The number of valid transactions far outnumber fraudulent ones. Also, the transaction patterns often change their statistical properties over the course of time. These are not the only challenges in the implementation of a real-world fraud detection system, however. In real world examples, the massive stream of payment requests is quickly scanned by automatic tools that determine which transactions to authorize. Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent.

II. ENSEMBLE CLASSIFICATION

Group learning systems rather produce different models. Given another model, the outfit passes it to all of its various base models, procures their gauges, and thereafter combines them in some fitting way (e.g., averaging or projecting a voting form). The majority of gathering learning procedures are regular, fitting across sweeping classes of model sorts and learning endeavours [6]. Social occasion learning is an effective method that has dynamically been embraced to combine diverse learning computations to additionally create overall assumption exactness [1][2]. Conceivably the most powerful spaces of investigation in coordinated AI have been to peruse systems for building incredible groups of understudies. The central disclosure is that outfits are habitually extensively more precise than the individual understudies that make them up [5].

When arranging an outfit learning method, just as picking the system by which to accomplish assortment in the base models and picking the joining procedure, one requirement to pick the sort of base model and base model learning computation to use. The combining strategy may restrict the sorts of base models that can be used.

III. METHODOLOGY

Perhaps the most dynamic spaces of exploration in administered AI have been to read techniques for building great gatherings of ensemble classification.

3.1 AdaBoost

AdaBoost represents Adaptive Boosting, its anything but a notable, successful procedure for expanding the precision of learning calculations. The arrangement of base classifiers, delivered by AdaBoost from the preparation set, is applied to the approval set, making an altered arrangement of loads. The preparation and approval sets are exchanged, and a subsequent pass is performed. Re-weighting and resampling are two techniques executed in AdaBoost[2]. The fixed preparing test size and preparing models are re-inspected by a likelihood appropriation utilized in every cycle. In term of re-weighting, all preparation models with loads allotted to every model are utilized in every emphasis to prepare the base classifier [9].

3.2 XGBoost Algorithm

The concept of boosting came to limelight when it was examined whether a "weak learner" could be made a "better learner" by using some kind of modifications. From statistics point of view, this process was similar to creating a "good hypothesis" from a relatively "poor hypothesis". According to Jason Brownlee, author of, a poor learner or a "weak hypothesis" is a model whose performance is slightly better than random chance. Hypothesis boosting involves the idea of filtering the observations [4]. Those observations which the weak learner can handle is left as it is and those observations that the weak learner cannot handle are focused on. In the AdaBoost algorithm, the weak learners were given more weights and the strong learners were given less weights and this weight was changed repeatedly until a proper model was found which could correctly classify the given samples. When a prediction was needed to be done, the majority vote of the weakest learners' prediction was taken and the corresponding weight was chosen for the weight of the ultimate prediction. One benefit of the gradient boosting model is that for different loss functions, new algorithms are not required to be derived; it is enough that a suitable loss function be chosen and then incorporated with the gradient boosting framework. Second, a weak learner is created to make the predictions. In gradient boosting a decision tree is chosen as a weak learner. Specifically, regression trees are used that produces real value output for splits and whose output can be added together, allowing subsequent outputs of different models to be added. This approach enables the improvement of the residuals in the predictions leading to more precise predictions. The trees are created in a greedy manner and often certain constraints are imposed in order to ensure that the weak learners continue to be weak learners and still the trees can be created using a greedy approach. Third, creation of an additive model to add up the predictions of the weak learners so as to reduce the loss function. This process of adding the trees happens one at a time. The output produced in the new tree is then added to the output of the pre-existing sequence of trees in order to improve the final output of the model. This process stops once the proper optimized value for the loss function is reached. XGBoost also has gradient boosting at its core [7]. However, the difference between simple gradient boosting algorithm and XGBoost algorithm is that unlike in gradient boosting, the process of addition of the weak learners does not happen one after the other; it takes a multi-threaded approach whereby proper utilization of the CPU core of the machine are utilized, leading to greater speed and performance. Apart from that, there is sparse aware implementation which also involves automatic handling of missing data values, then block structure to support the parallelization of tree construction, and the process of continued training so that one can further boost an already fitted model on new data. It is to be noted that XGBoost has been seen to dominate structured or tabular datasets on classification and regression and predictive modeling problems [8].

IV. EXPERIMENTAL RESULTS

We have considered the Credit card fraud detection dataset from the Kaggle to assess performance of Ensemble classification. The examinations have been led by utilizing Python programming dialect. The Python Scikit-learn is a bundle for information order, relapse, grouping and representation. The Credit card fraud detection informational collection has 284807 lines and 30 attributes. The objective class contains two qualities: 0 and 1 where 0 speaks to No fraud and 1 speaks to fraud. In characterization issues how class names are conveyed. So, in this information there are two class names i.e., The No fraud class has 284315 occasions and fraud class has 492 examples. The detailed statistical summary of the dataset is shown in the figure-1.

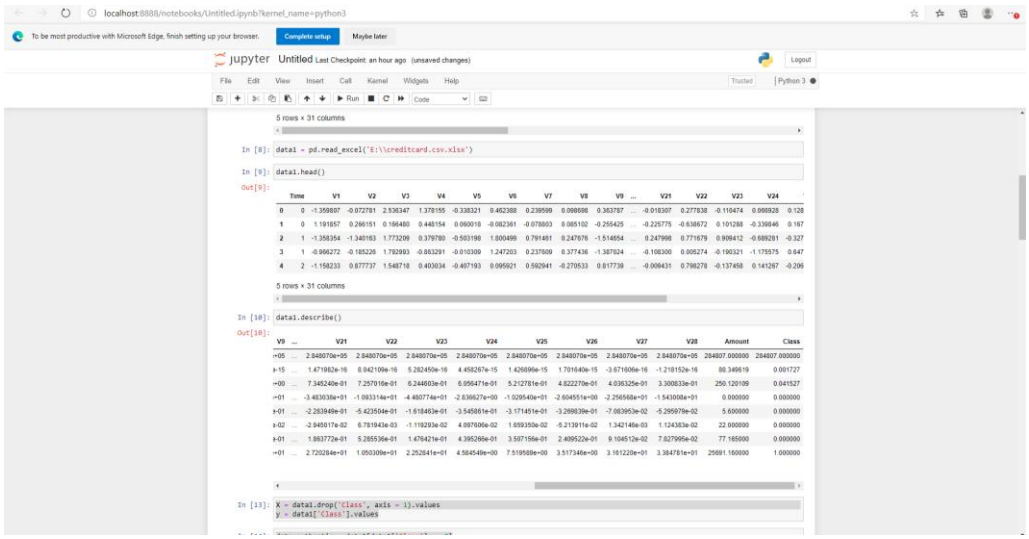

FIGURE-1: Detailed statistical summary of the dataset

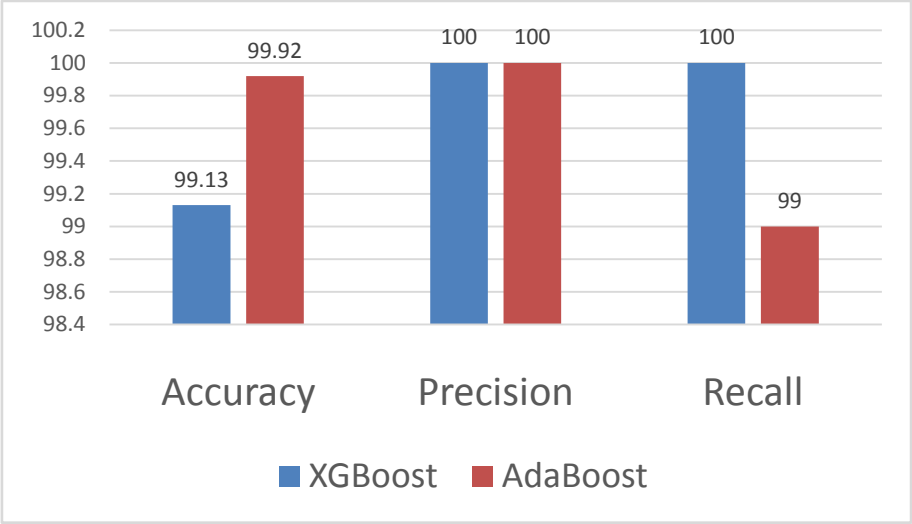

FIGURE-2: Results of classification algorithms

From the figure-2, we notice the exhibition of classification for XGBoost 99.13% of Accuracy and the AdaBoost has achieved the accuracy of 99.92%. So, the both algorithms have got highest accuracy, but only 0.79% difference of AdaBoost when compared to XGBoost. The screen shots of experimental results are shown in the figure-3 and figure-4.

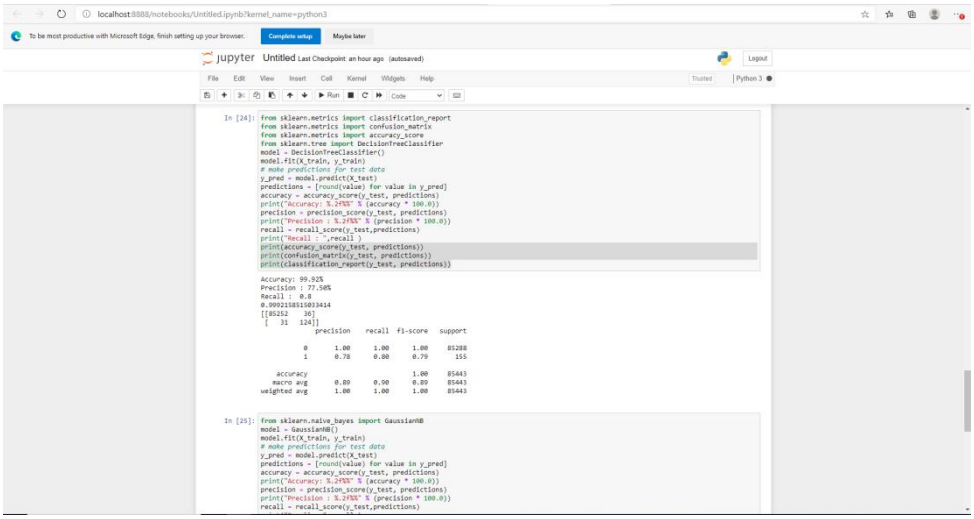

FIGURE-3: Screen shot of the Experimental results

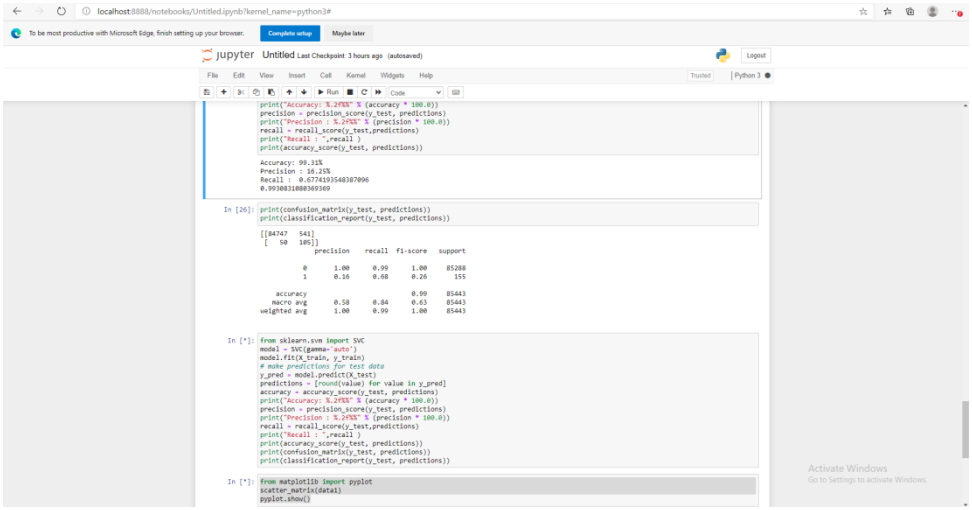

FIGURE-4: Screen shot of the Experimental results

V. CONCLUSION

This paper presents an XGBoost and AdaBoost-based financial system to detect transaction fraud. The results thus obtained show that the highest precision and accuracy of AdaBoost and XGBoost is 99 percent for credit card fraud detection problems. So, the AdaBoost and XGBoost models are options to recognize fraudulent credit card transactions precisely.

REFERENCES

[1] Dietterich TG, Ensemble methods in machine learning. In: Proceedings of Multiple Classifier System‖, vol. 1857. Springer; 2000. pp. 1–15.

[2] Freund, Y., and Schapire, R. E., ―A decision-theoretic generalization of on-line learning and an application to Boosting, J. Comput. Syst. Sci. 55(1):119–139, 1997

[3] Friedman, J. H. (2002). Stochastic gradient boosting. Computational statistics & data analysis, 38(4), 367-378.

[4] GanglongDuan, Xin Ma, "A Coupon Usage Prediction Algo Based OnXGBoost", 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 978-1-5386-8097-1/18/$31.00 ©2018 IEEE.

[5] G. Ravi Kumar, Venkata Sheshanna Kongara & Dr. G. A. Ramachandra, "An Efficient Ensemble Based Classification Techniques for Medical Diagnosis", International Journal of Latest Technology in Engineering, Management and Applied Sciences, Volume II, Issue VIII, Pages: 5-9, ISSN-2278-2540, August-2013

[6] Han J and Kamber M, Data Mining Concepts and Techniques. Morgan Kanufmann, 2006.

[7] Huiting Zheng, Jiabin Yuan, & Long Chen. Short-term load forecasting using emd-lstm neural networks with an XGboost algo for feature importance evaluation. Energies, 10(8):1168, 2017.

[8] Iyad LahsenCherif, AbdesselemKortebi, "On using eXtreme Gradient Boosting (XGBoost) Machine Learning algo for Home Network Traffic Classification", 978-1-7281-0117-0/19/$31.00 ©2019 IEEE.

[9] Zhang, C. X., Zhang, J. S., and Zhang, G. Y., ―An efficient modified Boosting method for solving classification problems‖. J. Comput. Appl. Math. 214(2):381– 392, 2008.