# Using Adaboost AND Majority Voting Techniques For Detecting Fradulent Transcations In Credit Card

## S Sharath[1*], Sherwin Sampath Doddamani[2], Lakshmikantha S[3]

[1,2,3]Department of Computer Science, East West Institute of Technology, Bengaluru, India

*Abstract* **:** Credit card fraud is a serious problem in financial services. Billions of dollars are lost due to credit card fraud every year. There is a lack of research studies on analyzing real-world credit card data owing to con dentiality issues. Nowadays digitalization gaining popularity because of seamless, easy and convenience use of e-commerce. It became very rampant and easy mode of payment. The experimental results positively indicate that the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards. Inspired by the recent novel idea of Trerngad [1], we also quantize the released gradients to ternary levels $\{-B, 0, B\}$, where B is the bound of gradient clipping. Voting based prediction aggregation provides the final predictions. A hybrid technique of under-sampling and oversampling is carried out on the skewed data. Capsule Network (CapsNet) is adopted to further dig some deep features on the base of the expanded features, and then a fraud detection model is trained to identify if a transaction is legal or fraud.

*Keywords*— AdaBoost, classification, credit card, fraud detection, predictive modelling, voting.

## I. INTRODUCTION

The credit card fraud detection as become a major cause for the loss of lot of money in financial and a business process and even in the enterprise. there are many techniques ie majority voting decision tree etc.. now a day's hackers are so brilliant that they are not able to identify is it a fraud or a normal transaction. in 2014 the us as lost v1.6 billion dollars due to the credit card fraudent .They can respond to the rise in credit card fraud has a big impact on the respond to it but they don't because if they do it they financial industry there has been an tremendous might definitely get caught so that well-educated Increase in electronic transactions during the recent fraudsters are so clever that due to these online Last decades[3]. Most of them make use of the online transaction methodology they can easily hack and Payment for shopping purposes or at the marketed fraudsters can do their job easily without getting This leads to the increase in online transactions into trouble. Nowadays these types of banking crime Using credit and debit cards evolving to the world are getting very common these days money gets Of effortless expenditure. In the visa transaction hefted and by the cyber-crime cannot be found easily Master card accreditation to web transactions to how and where the money got dispatched and all, so Furthermore physical card for logged of transaction Nowadays everybody gets scared to use online banking May be utilized. In day-to-day usage of credit card transactions or by using net banking for various purpose.

## II. RELATED WORK

A. Classification:
Classification is the normal and one of the easiest and common techniques used to solve the fraud . To differentiate various category of object. A model is use called as classification. Classification identifies the labels of object; labels are defined before , unarranged and far. The application like, detection of credit card fraud, healthcare fraud, automobile insurance, corporate fraud, etc. where classification become very useful to detect fraud.

B. Clustering:
Clustering is one of the unsupervised method were it has been labled as dots where there is x and y axis and the dots have been present inside the axis and which are far from one another with different colours

C.Prediction:
Prediction is used to predict the continuous value. Based on the historic data, it makes patterns, estimate the numeric, and ordered value for future. As per the study author stated that, predicted values are continuous value rather than discrete value.

D.Outlier Detection:
Outlier is a method where a data is different from the other data sets in a given point and which is used to differentiate from one another[2]. As per the study in paper, author stated that, "Data points that having different characteristics with compare to reminder of whole dataset are known as outlier". Hence, the authors mentioned that outlier detection is very critical issue in the field of credit card and data mining.

**E. Regression:**

Regression shows the relationship between more than one dependent and independent variable. Regression is one of the best statistical method. Hence, for several empirical studies regression is like a benchmark. Real-world data sets related to telecommunications fraud, computer network intrusion, and credit card fraud were evaluated. The results were displayed with visual appeal to data analysts as well as non-experts, as high-dimensional data samples were projected in a simple 2-dimensional space using the SOM.

## III.   METHODOLOGY

### A. MAJORITY VOTING

Majority voting is frequently used in data classification, which involves a combined model with at least two algorithms. Each algorithm makes its own prediction for every test sample. The final output is for the one that receives most of the votes, as follows.

Consider $K$ target classes (or labels), with $C_i$; $8i$ 2 3 D f1; 2; : : : ; $K$ g represents the $i$-th target class predicted by a classier. Given an input $x$, each classier provides a pre-diction with respect to the target class, yielding a total of $K$ prediction, i.e., $P_1$; : : : ; $P_K$ . Majority voting aims to produce a combined prediction for input $x$, $P$ .$x$/ D $j$; $j$ 2 3 from all the $K$ predictions, i.e., $p_k$ .$x$/ D $j_k$ ; $k$ D 1; : : : ; $K$ . A binary function can be used to represent the votes, *i.e.*,

$$k \quad 2 \quad i \quad D( \ 0; \quad \text{otherwise}$$

$$V \quad .\mathbf{x} \quad C/ \qquad 1; \quad \text{if } p_k \ .\mathbf{x}/ \ D \ i; i \ 2 \ 3 \qquad (1)$$

Then, sum the votes from all $K$ classier for each $C_i$, and the label that receives the highest vote is the final (combined) predicted class.

### B. ADABOOST

Adaptive Boosting or AdaBoost is used in conjunction with different types of algorithms to improve their performance. The outputs are combined by using a weighted sum, which represents the combined output of the boosted classier, i.e.,

$$\begin{array}{c} T \\ X \\ F_T \ .x/ \ D f_t\,(x) \qquad (2) \\ t\text{D}1 \end{array}$$

where every $f_t$ is a classier (weak learner) that returns the predicted class with respect to input $x$. Each weak learner gives an output prediction, $h(x_i)$, for every training sample. In every iteration $t$, the weak learner is chosen, and is allotted a coefficient, $_t$, so that the training error sum, $E_t$ , of the resulting $t$-stage boosted classier is minimized,

$$X$$

$$E_t \ D \qquad\qquad E \ [F_{t\ 1} \ .x_i/ \ C \qquad_t h(x_i)] \quad (3)$$

## IV.   RESULTS AND DISCUSSION

### BENCHMARK DATA

A publicly available data set is downloaded from the implementation and how it is performed . It contains a total of 284,807 transactions made in September 2013 by European cardholders. The data set contains 492 fraud transactions, which is highly imbalanced. Due to the con identicality issue, a total of 28 principal components based on transformation are provided. Only the time and the amount data are not transformed and are provided as such from our algorithms and experiments.
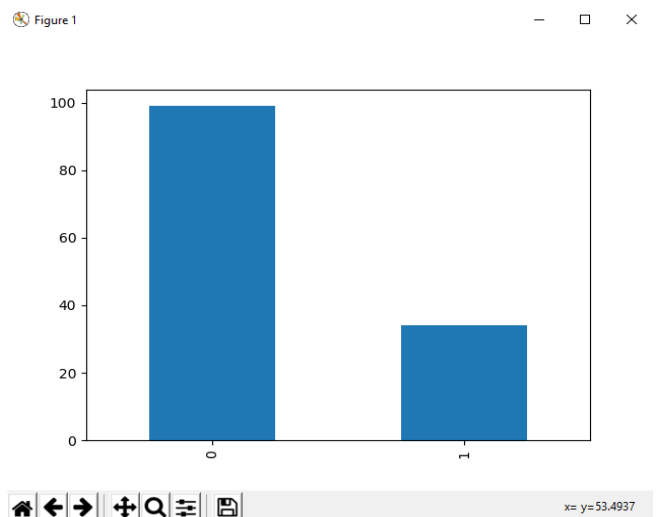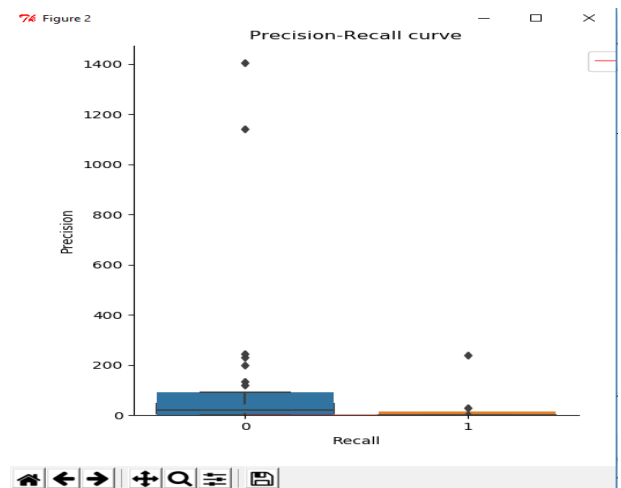


Figure1.Regression graph



Figure 2:Precision-Recall curve

## V.   CONCLUSION AND FUTURE SCOPE

Credit card fraud detection is a majority domain where it takes lot of time to find the fraudent[7]. From the paper we

come to know that adaboost has better accuracy, we analyzed machine learning is best in compare to prediction, clustering, outlier detection etc., Several standard models which include NB, SVM, and DL have been used in the basic evaluation. In addition, we upload ternary gradients instead of the exact gradients to reduce communication burden.

## VI.   FIELD WORK AND RESULT

we can improve the techniques by combining two or more algorithms together which should to lead to the providing alternatives results and finally we can choose the majority voted algorithm and their benefits by providing the upcoming equations and the results .we can improve the credit card fraud detection by alternating the digital and physical techniques which would  lead to providing better accuracy in future.

### A.Authors and affillation
S Sharath, B.E, Department of  Computer Science and Engineering, East West Institute of Technology.
Sherwin Sampath Doddamani, B.E, Department of Computer Science and Engineering, East West Institute of Technology.
LakshimiKantha S, B.E, MTech, Assistant Professor Department of  Computer Science and Engineering, East West Institute of Technology.

## ACKNOWLEDGEMENT

### REFERENCES

[1] N. Mahmoodi, E. Duman, "Detecting credit card fraud by Modified Fisher Discriminant Analysis", Elsevier Expert System with Application, 2015, pp. 2510-2516.
[2] Y. Sachin, E. Duman, "Detecting Credit Card Fraud by Decision Tree and Support Vector Machine", In Proceedings of the international multi Conference of Engineers and Computer Scientists, Hong Kong, 2011, pp. 1-6.
[3] V. Van Vlasselaer et al., "APATE : A novel approach for automated credit card transaction fraud detection using network-based extensions[J]," Decision Support Systems, 2015, 75:38-48.
[4] Mafarja M , Mirjalili S. Hybrid Whale Optimization Algorithm with simulated annealing for feature selection[J]. Neurocomputing, 2017, 260:302-312.
[5] Kecman, Vojislav; Learning and Soft Computing Support Vector Machines, Neural Networks, Fuzzy Logic Systems, The MIT Press, Cambridge, MA, 2001.
[6] Fadaei Noghani, F., and M. Moattar. "Ensemble Classification and Extended Feature Selection for Credit Card Fraud Detection." Journal of AI and Data Mining 5.2 (2017): 235-243.
[7] The Nilson Report. (2015). U.S. Credit & Debit Cards 2015. David Robertson.
[8] Stolfo, S., Fan, D. W., Lee, W., Prodromidis, A., & Chan, P. (1997). Credit card fraud detection using meta-learning: Issues and initial results. In AAAI-97 Workshop on Fraud Detection and Risk Management.