# A Credit card fraud detection using Naïve Bayes and Adaboost

Sushma, Dr Mary Cherian
1Student, Dr. Ambedkar Institute of Engineering and Technology, Bengaluru, India
2Professor, Department of CSE, Dr. Ambedkar Institute of Engineering and Technology, Bengaluru, India

## Abstract

Credit card fraud is a major issue in financial services. Billions of dollars are lost because of MasterCard misrepresentation consistently. There is a lack of research on analyzing real-world MasterCard information owning to privacy issues. In this paper, AI calculations are utilized to detect MasterCard misrepresentation. Standard models are first utilized. At that point, hybrid techniques which use AdaBoost and majority voting methods are connected. To assess the model efficiently, a freely accessible MasterCard informational collection is utilized. At that point, a real-world charge card informational collection from a financial organization is analyzed. Also, noise is added to the information tests to additionally survey the strength of the calculations. The experimental results exactly shows that, the majority voting technique accomplishes greater accuracy rates in detecting fraud cases in charge cards.

***KEY TERMS*** *AdaBoost; arrangement; MasterCard; predictive modelling;majority voting.*

## INTRODUCTION

Fraud is a cheating or a wrongful or criminal activity, aimed to increase financial or personal gain. To avoid loss from the fraud, there are two mechanisms: Fraud Prevention and Fraud detection. Fraud prevention mechanism is a most protective and proactive technique, it prevents fraud from starting. Then, second mechanism fraud detection is guessing the fraudster. This instrument is required to avoid fake exchange.Master card fraud is connected with illegal utilization of MasterCard data for buyers. The MasterCard transaction can be established physically or digitally. In physical transactions, Master card is utilized duringtransactions. In digital transactions, this can happen over the phone or web. The cardholders essentially gives card number, expiry date, and card check number by means of phone or website.With the rise of e-commerce in previous years, the utilization of MasterCard has increased rapidly. The number of MasterCard transactions in the year of 2011 in Malaysia was about 320 million, which increased to about 360 million in the year of 2015.Due to the rise of MasterCard utilization, the number of fraud cases have been consistently increased so it is most effective problem in the real-world. While numerical authentication methods are applied in this MasterCard fraud cases, but this method is not most detected one, because the fraudsters hide their details like identity and location in the internet. This problem has a big impact on financial industry also. This MasterCard fraud problem affects both admin and user. It affects card issuer fees, charges and administrative charges. So the merchants need to bear the loss, some goods are priced higher or discounts are reduced.Therefore, it is important to reduce theloss,and an effective fraud detection system to reduce or eliminate fraud cases is important. There have been various studies on MasterCard fraud detection. Machine learning and related techniques are most commonly utilized, which includesartificial neural networks, Naive Bayes, Decision Trees, logistic regression and support vector machine (SVM). The above models combined by several methods together to form hybrid model. The AdaBoost and majority voting are applied to recognize the MasterCard misrepresentation.

## RELATED STUDIES

In this segment, single and hybrid AI calculations for financial applications are explored. Different financial applications from MasterCard fraud to financial report fraud are investigated.

### Single Methods

For MasterCard fraud detection, Random Forest, Support Vector Machine and Logistic Regression were inspected in [5]. The informational collection comprised of one-year transactions. Information under-testing was utilized to analyze the algorithms implementation, with Random Forest showing a superior execution has contrasted withSVM and LOR [5]. An Artificial Immune Recognition System (AIRS) for MasterCard misrepresentationwas proposed in [6].AIRS is an improvement over the standard AIRS model, where negative choice was used to accomplish higher exactness. This brought about an expansion of precision by 25.0% and diminished framework reaction time by 40.0% [6].

A MasterCard fraud detection framework is proposed in [9], which comprised of standard based channel, exchange history database, and Bayesian student. The Dumpster–Shafer hypothesis consolidated different evidential data and made an initial belief, which was utilized to order a transaction as ordinary, suspicious, or abnormal .If a transaction was suspicious, thebelief further evaluated utilizing exchange history from Bayesian learning [9]. Recreation results are showed 98.0% true positive rate [9]. The adjusted Fisher Discriminant work was utilized in Master Card misrepresentation identification in [8].

### Hybrid Models

Hybrid models are combination of multiple individual models. A hybrid model consisting of the Multilayer Perceptron (MLP) neural network, SVM, LOR, and Harmony Search (HS) optimization was utilized in [10] to detect corporate tax evasion. HS was fulfilling for finding the best parameters for the classification models. Utilizing information from the food and textile sectors in Iran, the MLP with HS optimization acquired the highest accuracy rates at 90.07% [11]. A hybrid clustering system with outlier detection capability was utilized in [12] to detect fraud in lottery &online games. The system aggregated online algorithms with statistical information from the input information to identify a number of fraud types. The training information set was compressed into the main memory while new data samples could be incrementally added into the stored data cubes. The system achieved a high detection rate at 98%, with a 0.1% false alarm rate [12].

## PROPOSED SYSTEM

In this paper, a sum of twelve AI calculations is utilized for detecting MasterCard fraud. The algorithms were extended between standard neural systems and profound learning models. They are evaluated using both benchmark and real-world MasterCardinformation sets. In addition, the AdaBoost and majority voting methods are applied for forming hybrid models. To further evaluate the robustness and reliability of the models, noise is added to the real-world data set. The key contribution of this paper is the evaluation of a variety of machine learning models with a real-world credit card data set for fraud detection. While other researchers have used various methods on publicly available data sets, the data set used in this paper is extracted from actual credit card transaction information over three months.

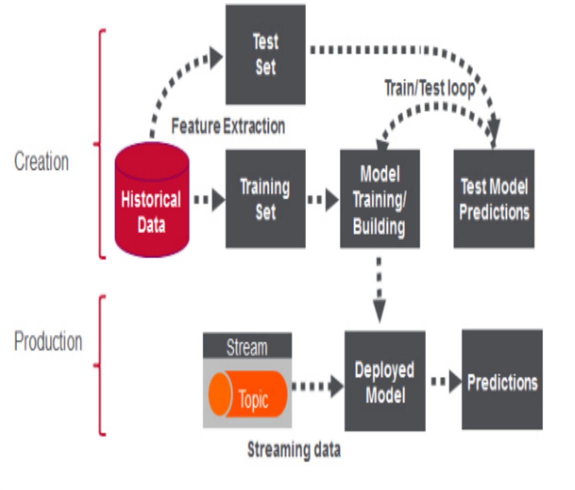


Fig 1: System Architecture

### Machine Learning Algorithms

A total of twelve algorithms are used in this experimental study. They are used in conjunction with the AdaBoost and majority voting methods. The details are as follows.

### Algorithms

Decision Tree: To detect the financial fraud and to minimize the fraud, but not fully decrease the fraud.

Random Tree: Itworks as a DT administrator, just an arbitrary subclass of highlights is accessible. It learns from both nominal and numerical information samples. The subclass size is defined using a subclass ratio parameter.

Random Forest: The user sets the number of trees. The resulting model employs voting of all created trees to determine the final classification outcome.

Naïve Bayes: Naïve Bayes uses the Bayes' theorem with strong or naïve independence assumptions for classification. Certain features of a class are assumed to be not correlated to others. It requires only a small training data set for estimating the means and variances is needed for classification.

Gradient Boosted Tree: The Gradient Boosted Tree is an ensemble of classification or regression models. It uses forward-learning ensemble models, which obtain predictive results using gradually improved estimations. Boosting helps improve the tree accuracy.

SVMReduction: To reduce cumulative misclassification cost .but significantly reduces the misclassification cost.

NN: Find out fraud transactions and only find out the fraud transaction, not reduce the fraud.

## AdaBoost and Majority Voting

Adaptive Boosting or Ada Boost is utilized in conjunction with different types of algorithms to improve their performance. The outputs are combined by utilizing a weighted sum, which represents the combined output of the boosted classifier. AdaBoost tweaks weak learners in favor of misclassified information samples. It is, however, sensitive to noise and outliers. As long as the classifier performance is not random, AdaBoost is able to improve the individual results from different algorithms.AdaBoost helps improve the fraud detection rates, with a noticeable difference for NB, DT and RT, which produces a perfect accuracy rate. The most significant improvement is achieved by LIR. 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 the majority of the votes.The majority voting method achieves good accuracy rates in detecting fraud cases in MasterCard.

## EXPERIMENTS

### *Investigational Setup*

In thisTrial setup, the MasterCard informational index and quantity of false transactions is an exceptionally little as contrasted and the absolute number of exchanges are made. With a slanted informational index, the subsequent exactness does not present on a precise portrayal of the framework execution. Misclassifying a genuine exchange causes poor client administrations, andneglecting to identify extortion cases makes misuse the money related foundations and clients.This information unevenness issue causes execution issues in AI calculations .The class with the majority samples influences the results.

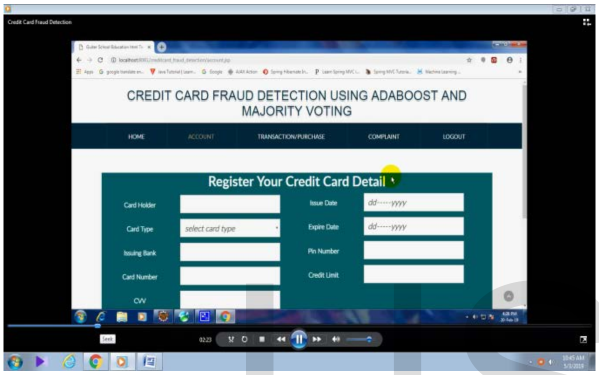
### *Benchmark Data*

A freely accessible informational index is installed from [3]. It providesa sum of 284,807 exchanges in a year 2013(September) published by European cardholders. The informational collection contains 492 misrepresentation exchanges, which is highly imbalanced. Because of the classification issue, an aggregate overall 28 essential segments dependent on change is given.

### *Certifiable Information*

The genuine MasterCardinformational index from a money related foundation at Malaysia is utilized in test. It depends on cardholders from the South-East Asia region in the year of 2017.Sum of 287,224 exchanges are recorded, with 102 of them delegated fraud cases. The information comprise of a period arrangement of exchanges.

## EXPERIMENTAL RESULT



Credit Card Fraud Detection

CREDIT CARD FRAUD DETECTION USING ADABOOST AND MAJORITY VOTING

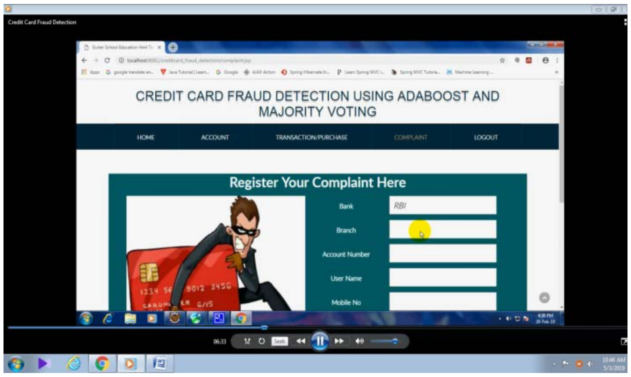
HOME ACCOUNT TRANSACTION/PURCHASE COMPLAINT LOGOUT

Register Your Credit Card Detail

| | | | |
|---|---|---|---|
| Card Holder | | Issue Date | dd-----yyyy |
| Card Type | select card type | Expire Date | dd-----yyyy |
| Issuing Bank | | Pin Number | |
| Card Number | | Credit Limit | |
| CVV | | | |

**Fig2: Registration Card Details in Credit Card**

The MasterCard is most ordinary way for go to line of credit. Similarly, it is given by a bank or economic favor. The user can enter the card details in form indicted in fig.2.



Credit Card Fraud Detection

CREDIT CARD FRAUD DETECTION USING ADABOOST AND MAJORITY VOTING

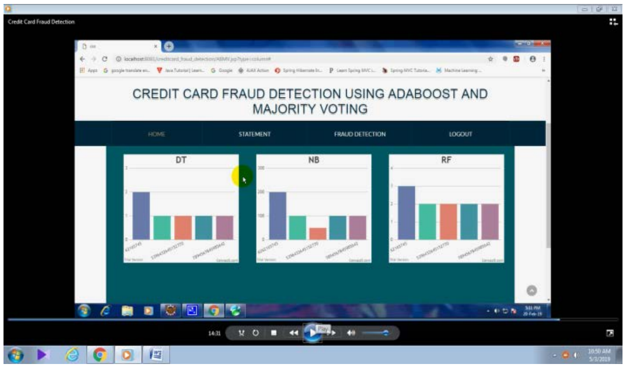
HOME ACCOUNT TRANSACTION/PURCHASE COMPLAINT LOGOUT

Register Your Complaint Here

| | |
|---|---|
| Bank | RBI |
| Branch | |
| Account Number | |
| User Name | |
| Mobile No | |

**Fig3: Registration MasterCard complaint about theft.**

Lodge a complaint about theft. Nowadays a person’s financial account details can be fetched easily due to which credit card frauds have been increased.Hence forth a user can file a complaint with bank to block the card or the account as indicated in fig3.



Credit Card Fraud Detection

CREDIT CARD FRAUD DETECTION USING ADABOOST AND MAJORITY VOTING

HOME STATEMENT FRAUD DETECTION LOGOUT

**Fig4: Graphical representation of MasterCard Fraud using Adaboost and Majority Voting.**

The above graph represents the rate of fraud occurred by online purchase using Majority Voting. And the methods used to detect frauds are: Decision Tree, Naïve Bayes and Random Forest. X-axis represents the different methods and Y-axis represents the year. Dark Blue represents the maximum theft occurrence and Red represents the average theft, whereas Green and Blue represent minimum theft. Finally Purple represents the overall theft occurred throughout the year.

## CONCLUSIONS

Various best machine learning algorithms are used for MasterCard fraud. In this proposed system, proposed models NB, SVM, and DL are utilized in the experimental assessment.An openly accessible charge card informational index has been utilized for assessment utilizing singular (standard) models and half & half models utilizing AdaBoost and majority voting.The MCC (Matthews Correlation Coefficient) measurements computes the presentation measures,as it takes into account the true and false positive and negative predicted outcomes.The best MCC score larger part, accomplished utilizing dominant part casting a ballot. A genuine Credit card informational index from a monetary organization has additionally been utilized for assessment. A similar individual and half breed models have been utilized. An ideal MCC score is obtained by utilizing AdaBoost and lion's share casting a ballot techniques, since that blend technique is shown in solid execution. This proposed idea has improved web based learning models. Utilize the web guidance to permit the speedy consciousness of master card fraud. To further evaluate the hybrid models, noise from 10% to 30% has been added into the data samples. The majority voting method has yielded the bestMCCscore of 0.942 for 30% noise added to the data set. This shows that the majority voting method offers robust performance in the presence of noise.

## REFERENCES

[1] *S. Bulkan, Y. Sahin, and E. Duman* published "A cost-sensitive decision tree approach for fraud detection," *Expert Systems with Applications, vol. 40, no. 15, pp. 5916–5923, in 2013.*

[2] *A. O. Adewumi and A. A. Akinyelu, published* "A survey of machine-learning and nature-inspired based credit card fraud detection techniques," *International Journal of System Assurance Engineering and Management, vol. 8, pp. 937–953, in 2017.*

[3] *A. Srivastava, A. Kundu, S. Sural, A. Majumdar, published* "Credit card fraud detection using hidden `Markov model," *in IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, in 2008.*

[4] *J. T. Quah, and M. Sriganesh published*"Real-time credit card fraud detection using computational intelligence," *Expert Systems with Applications, vol. 35, no. 4, pp. 1721–1732, in 2008.*

[5] *N. S. Halvaiee and M. K. Akbari published*"A novel model for credit card fraud detection using Artificial Immune Systems," *Applied Soft Computing, vol. 24, pp. 40–49, in 2014.*

[6] N. Mahmoudi asnd E. Duman, "Detecting credit card fraud by modified Fisher discriminant analysis," *Expert Systems with Applications, vol. 42, no. 5, pp. 2510–2516, in 2015.*

[7] *S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar*, "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning,*" Information Fusion, vol. 10, no. 4, pp. 354–363, in 2009.*

[8] *E. Duman and M. H. Ozcelik*, "Detecting credit card fraud by genetic algorithm and scatter search," *Expert Systems with Applications, vol. 38, no. 10, pp. 13057–13063, in 2011.*

[9*] P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose,* "Detection of financial statement fraud and feature selection using data

mining techniques,” *Decision Support Systems, vol. 50, no. 2, pp. 491–500, in 2011.*