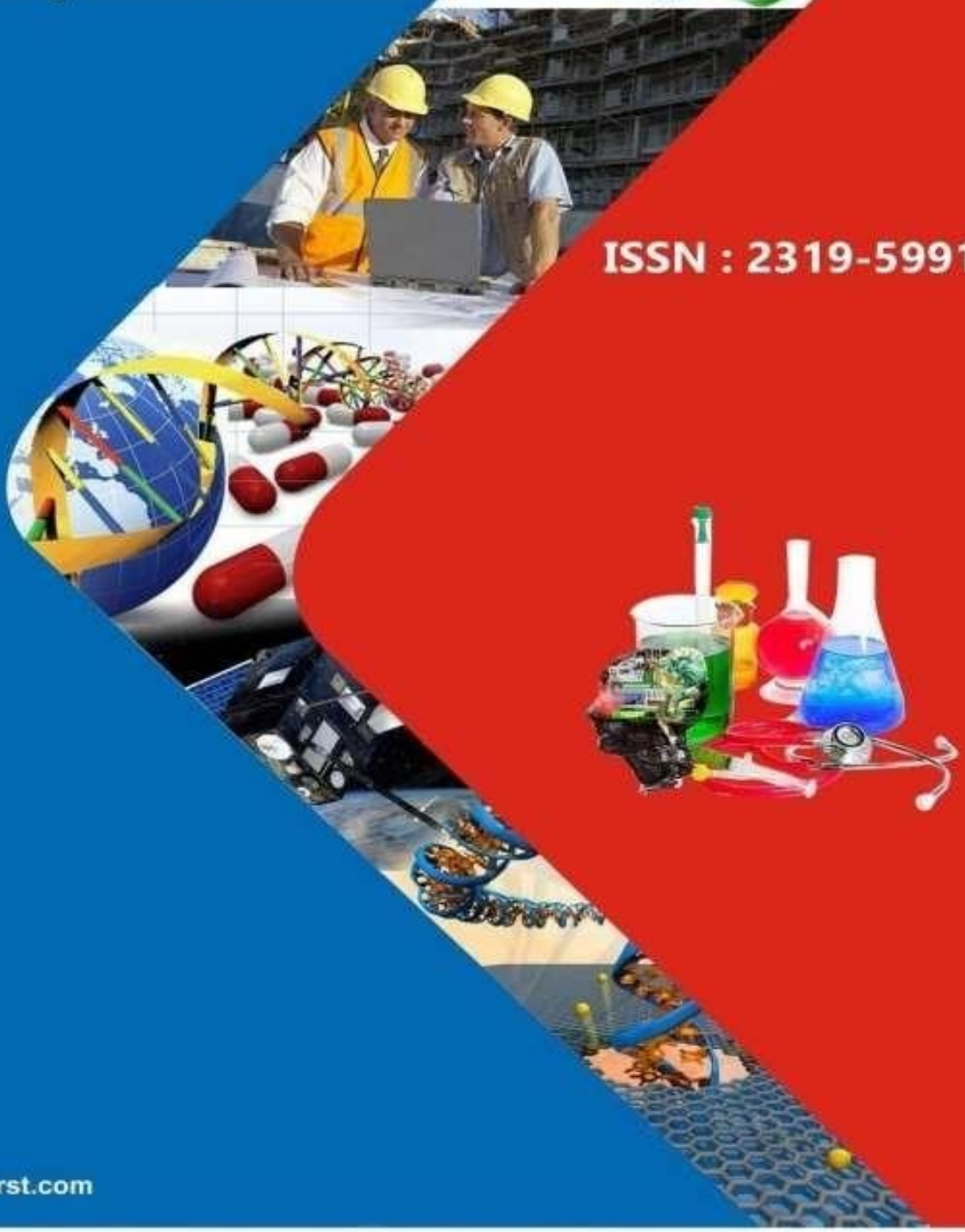
International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991



www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

USING ADABOOST AND MAJORITY VOTING TO IDENTIFY CREDIT CARD FRAUDULENT ACTIVITY

C. Yosepu, D. Sai Kiran, K. Rammohan, G. Ganapathy Babu
Assistant Professor Department of Computer Science and Engineering,
St. Martin's Engineering College, Secunderabad, Telangana, India

ABSTRACT

Typically, credit card fraud occurs when a card is taken for an unlawful use or even when the fraudster exploits the credit card details for personal gain. There are a lot of credit card issues in the globe nowadays. The credit card fraud detection system was created in order to identify fraudulent activity. The primary goal of this research is to concentrate on machine learning techniques. The Adaboost algorithm and the random forest method are the algorithms that are employed. The accuracy, precision, recall, and F1-score of the two algorithms are used to determine their outcomes. The confusion matrix is used to plot the ROC curve. After comparing the Random Forest and Adaboost algorithms, the optimum algorithm for detecting fraud is determined by evaluating its accuracy, precision, recall, and F1-score.

Keywords: Credit card fraud, fraudulent activities, Random Forest, Adaboost

I. INTRODUCTION

A mastercard could be a card which allows people to shop for items without cash. once they buy something, a sales clerk uses it to charge the money needed to their account, therefore the person pays later. It provide you with a line of credit that will be accustomed make purchases, balance transfers and/or cash advances and requiring that you just pay back the loan amount within the future. When using a mastercard, you may have to make a minimum payment of each month by the date on the balance.

A master card consists of a magnetic strip. The secondary tracks and primary tracks data are store with encoding in the magnetic strip which contains the information about customer's master card number with is complete name and its master card expiry date with country code. Additional information is stored within the third track. Fraud is typically performed by someone or thing intended to deceive others.

Frauds can be done with help of the waste and abuse also include the fraud payment, hiding, illegal financing, accessing security ,cyber security within the past, organizations had to require a fraud prevention and fraud detection techniques which uses business rules and analytics for checking the anomalies and to make alert for different data sets.

This maybe stepped off by preventing and detecting the fraud yet to be occurred or has been occurred. Prevention of mastercard frauds deals with stopping the frauds to occur. While Detection of mastercard is finished to predict

the frauds that has been occurred or possible to occur.Credit card fraud was ranked the primary quite Identity theft fraud - accounting for 35.4 percent of all identity theft fraud in 2018.In 2018, unauthorized financial fraud losses across payment cards and remote banking totaled £844.8 million within the UK. Whereas banks and card companies prevented £1.66 billion in unauthorized fraud in 2018.[3]

II. AIMS AND OBJECTIVES

a) AIM

The objective of project is to minimize the master card frauds. This setup lessens the masquerade by checking on the transactions moreover as funds of individuals. Most of the population uses the net and public platforms moreover as modes of communication to share their day to day activities which has their personal feelings. This sort of information is sufficient for the fraudsters to form fraudulent activities. The utilization of this setup helps cybernetics and web security to acknowledge the thieves.[3]

b) OBJECTIVES

The objective of the project is to make an authentic, genuine, precise and handy orderliness setup. The objectives of the system are reducing the time and value of confirming the transactions and withdrawals from the shoppers as this might result in the discontentment.

III. LITERATURE SURVEY

It is very necessary to check and try to solve the credit card fraud so that it becomes safe for the other entire user

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

to do online transaction all over the world without worrying about fraud during the transaction. For that the use of different techniques used to detect and solve online credit card fraud.

Paper1: Credit card fraud detection using Bayesian and neural networks:

The paper discuss about how to detect the online master card fraud using the Bayesian and neural network. It was used during the era of digitalization to reduce online master card fraud detection on ecommerce website. Now digitalization is mostly used in this era in which master card are mostly used so that's why master card fraud detection is given high importance in the society of this eraIt use the machine learning algorithm for detecting the fraud.[1]

Paper2: Credit card fraud detection using hidden Markov model:

Due to increase in advancement of electronic commerce technology the use of master card has increased this result in increase of master card fraud .Thus the use of most drastic method to solve this problem is hidden markov model. This model obtains high fraud coverage area combining with the low false alarm. [2]

Paper3: Credit card fraud detection using the auto- encoder and restricted Boltzmann machine base on the deep learning model:The hacker uses different types of method for fraud in online transaction and those methods always changes .This fraud does not have specific pattern so for detecting those fraud uses the different method .The most useful method during this situation is deep learning based algorithm. This automatically detects the fraud using the Boltzmann machine.[4]

IV. EXISTING SYSTEM

Three methods to represent frauds are mentioned here. The primary method is Clustering Model(CM). CM is employed to classify valid unlawful, criminal withdrawals transaction using data clusterization of regions of parameter value. K-means algorithm is simplest method to implement the fraud detection method with the approximate solution and k-means algorithm contains popularity for being simple, easy to access, good convergence speed and adaptability for data. The Second method used is Gaussian mixture model. This model is employed to imitate the chance of the mastercard user's earlier actions in order that the chance of recent actions will be measured to spot any deformity from earlier actions. Bayesian Networks are accustomed illustrate the entropy of various fraudulent measures. Validation of mastercard number is finished by combining Luhn Algorithm along with the K-Means Algorithm.[1]

V.COMPARATIVE STUDY:

Sr. No	Paper Name	Author/ publication	Technology	Advantage	Disadvantage
1.	Fraud detection of credit card using adaboost and majiority voting [3]	Randhawa, Kuldeep	AdaBoost and majority voting.	It reduces the master card fraud and predicts the future frauds.	
2.	Fraud detection of credit card using hidden markovmodel [2]	Srivastava, Abhinav	Hidden Markov Model	HMM can detect the fraudulent activity at the time of the transaction	HMM cannot detect fraud with a few transactions
3.	Fraud detection of credit card using deep learning based on auto-encoder and restricted boltzman machine [4]	Apapan Pumsirirat, Liu YanSchool	Deep Learning is based on the use of Auto- Encoder and Restricted Boltzmann Machine	A key advantage of deep learning is the analysis and learning of a massive amount of unsupervised data. It can extract complex patterns	Recognization of image is done using deep learning. For explain their other domain no information is available. All data for algorithm are not available to cover all the library of deep learning algorithm.
4.	Credit Card Fraud Detection Using Random Forest.[5]	Bhattacharyya, Siddhart, et al	Random Forest	It checks importance of variables in regression classification	collative comparison measure that reasonably represents the gains and losses due to fraud detection is proposed

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

VI. PROBLEM STATEMENT

It uses the knowledge of the clothed to be fraud for detection of fraud in master card while transaction of the money during the online payment on e-commerce website. This model is then wont to identify whether a replacement transaction is fraudulent or not. Aim is to try to detect almost 90% of the fraudulent transactions while minimizing the wrong fraud classifications

VII. PROPOSED SYSTEM

The master card fraud detection was proposed which is consisted of the rule-based filter. The filter consists of the Dumpster-Shafer, transaction history and simple Bayesian learner. The theory of Dumpster-Shafer combines various evidential information and initial belief which has been created. This is used to classify the transaction as normal, suspicious, or abnormal. If a transaction was suspicious, the assumption was further evaluated using transaction history from Bayesian learning. The results indicate the 98% true result simultaneously. The modification made the normal functions to become more sensitive to big instances.The variances were calculated using the weighted average which allows the learning of transactions. The CDFM model managed to distinguish fraudulent filings from non-fraudulent ones. [3]

VIII. ALGORITHM

1) NAIVE BAYESIAN ALGORITHM**:**

Algorithm 1: Training Algorithm
Step 1: M= Set of Documents
Step 2: N= Class of document positive, negative.
Step 3: V= Extract of features Vector (A) Step 4: D = Total Number of training doc Step 5: C= Distinguished class
Step 6: For c in C
Step 7: Dc= Number of documents with class c.
Step 8: Prior[c] = Dc/ D Step 9: For w in V do Step 10: Likelihood
[w][c]=(count(w,c)+k)/((k+1)* Dumber of words in class c)
Step 11: Return prior, likelihood

Algorithm 2: Testing Algorithm
Step 1: V = Extract Feature Vector(t)
Step 2: For k in K do //k-Value of likehood.
Step 3: score[k] = prior[k] Step 4: for w in V do
Step 5: score[k]=
score[k] * likelihood[w][c] Step 6: End second for loop.
Step 7: End first for loop.
Step 8: Return argmax(score[c])

2) RANDOM FOREST ALGORITHM:

Step 1: Start the data process
Step 2: Execute the data preprocessing.
Step 3: Develop the decision trees associated with the selected data points using RF Classifier as given below.
classifier= randomforestclassifier (n_Estimators= x, C ="entropy")
classifier.fit(X_train, Y_train)
Step 4: Predicting the test result
x.p = classifier. predict(train_data)
y.p = classifier. predict(test_data)
Step 5: Repeat first to steps again.
Step 6: Find the different prediction for different new data points and also different decision tree using latest data points, and assign them to the category that fits into the subsets.

3) DECISION TREE ALGORITHM:

Step 1: The root node is used at the start of Decision tree, say P that contain in the complete dataset.
Step 2: Use Attributes Selection Measure to find attribute in the dataset.
Step 3: Divide set into subset that contains values for attributes.
Step 4: Decision tree node is generated.
Step 5: Make new decision trees using the subsets of the dataset created in step -3. Continue process until a stage

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

is reached.

IX .MATHEMATICAL MODEL

1) MAJORITY VOTING:

- Majority voting is usually utilized in data classification, which involves a combined model with a minimum of two algorithms.
- Once the majority votes is received by that then it is called as the final output is as follows,
- Consider K target classes (or labels), with Ci={1,2,…,K} represents the i -th target class predicted by a classifier.
- For given input x, prediction is provided by each classifier with respect to the target class, yielding a total of K prediction, i.e., $P1,\ldots, .Pk$
- Majority voting aims to produce a combined prediction for input x , P(x)=j,j from all K predictions, i.e., $Pk\ (x) = jk$, k=1,…,K . A binary function are often wont to represent the votes, i.e.,

$V7(\mathbf{x}\in C") = \{1, ifPk(x) = i, i\mathcal{E}\hat{}$
0, $otherwise$

- Then, sum the votes from all K classifiers for each Ci , and the label that receives the highest vote is the final (combined) predicted class.[3]

2) ADABOOST:

- Adaptive Boosting or AdaBoost is employed in conjunction with differing types of algorithms to enhance their performance.
- The outputs are combined by using a weighted sum, which represents the combined output of the boosted classifier, i.e.,(x) = $\sum T$ 1 f (x)
- Where every ft is a classifier that returns the predicted class with respect to input x.
- Each weak learner gives an output prediction, h(Xi)

, for every training sample.

- In every itration t, it chooses the weak learner and allotscoefficient $\alpha t$and calcu training error sumEt for testing t-stage boosted classifier is

minimized,$E\ \ t$
t=$\sum i$ [ft−1(xi) +αth(xi)]

- Where Ft-1(x) is the boosted classifier built in the previous stage, E (F) is the error function, and (x)= αth(x) is weak learner taken into consideration for the final classifier.
- AdaBoost weak learners in favor of misclassified data 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.[3]

3) NAIVE BAYESIAN:

- The basis of Naive Bayes algorithm is theorem or alternatively referred to as Bayes' rule or Bayes' law.
- It gives us a technique to calculate the contingent probability, i.e., the probability of a happening supported previous knowledge available on the events.
- More formally, theorem is stated because the following equation:

4) RANDOM FOREST:

- Classification and regression is used in the Random Forest Algorithm.
- Summarily, it's a group of decision tree classifiers.
- Random forest has advantage over decision tree because it corrects the habit of overfitting to their training set.
- A subset of the training set is sampled randomly in order that to coach each individual tree then a choice tree is constructed, each node then splits on a feature selected from a random subset of the complete feature set.
- Even for giant data sets with many features and data instances training is extremely fast in random forest and since each tree is trained independently of the others.
- The Random Forest algorithm has been found to provides a decent estimate of the generalization error and to be proof against overfitting.

[5]

$xP\ (\ )\ .\ P(c)$
$P(c/x) =$ c
$P(x)$

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

P(c/x) = P(x1/c)*P(x2/c)* *P(xn)*P(c)

X. SYSTEM ARCHITECTURE

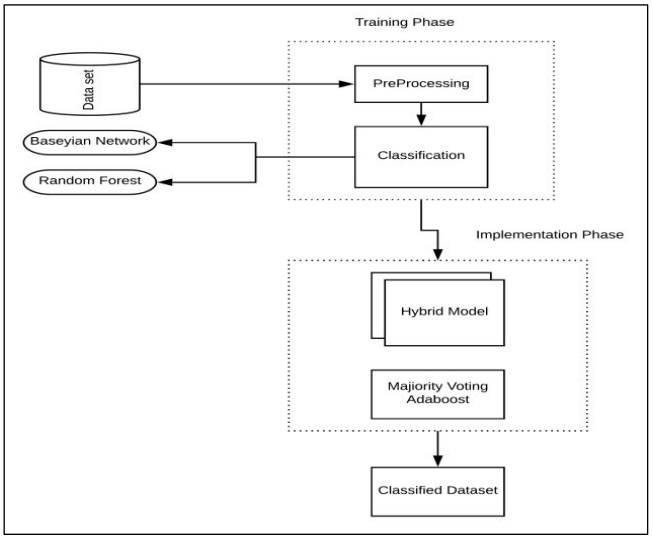


Fig.No.1.System Architecture

P(A\B) = P(B\A). P(A)

P(B)

- The statement is understood first then looked at the proof of the statement. The components of the above statement are shown in the following points which are part of the above formula:

a) P (A/B): Probability (conditional probability) of occurrence of event A given the event B is true or not in the given data.
b) P(A) and P(B): Probabilities of the occurrence of event A and B respectively.
c) P(B/A): On the basis of event A the probability of the event B is given.

- The terminology within the Bayesian method of probability (more commonly used) is as follows:

a) A is termed the proposition and B is termed the evidence.
b) P(A) is termed preceding probability of proposition and P(B) is termed preceding probability of evidence
c) P (A/B) is termed the posterior.[3]
d) $Posterior$ =
($Likelihood$). ($PropositionProbability$)

$Evidenceprobability$

Description:
The higher than design describes the work structure of the system. The system imports the important time information that contain the info of users and there mastercard range and every one the group action. Then the system creates knowledge the info the information set of every users consistent with his or her group action and it send all data set values to machine learning algorithmic rule.
The adaboost and majority voting are used to detect the errors occurred during the online transaction. The machine learning algorithmic rule processes all dataset and makes classification on basis of your time, location and transactions. The machine learning algorithmic rule then makes hybrid model of every group action. The hybrid model is combination of multiple individual models. Then system checks if there's any fraud in registered mastercard users by mistreatment adaboost and majority ballot. If there's any fraud then it creates the fraud detection table and tries to unravel the matter. The machine learning algorithms used are support vector machine and linear regression model. In these they are using two main machine learning algorithm Naïve Bayesian and the

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

random forest algorithm.[3]

XI. ADVANTAGES

1. The fraudster, with his new source of income will have sufficient revenue to enjoy his day to day life ,that may be better than most
2. As well as, bigger finance schemes.
3. Persons who are heavily indebted might go to “Credit Doctors”, where they’ll obtain stolen master cards they use to pay off their debts.
4. The persons investing their time in master card frauding schemes are usually those that are extremely intelligent.
5. This frauding medium gives them an chance to create on their skills and develop it to just right.

XII. DESIGN DETAILS

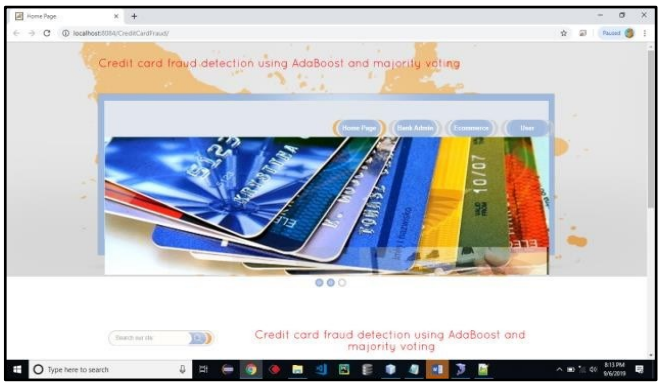


Fig.No.2. Home Page

It is the main home page of the credit card fraud detection. It contains the main login for user and the admin.

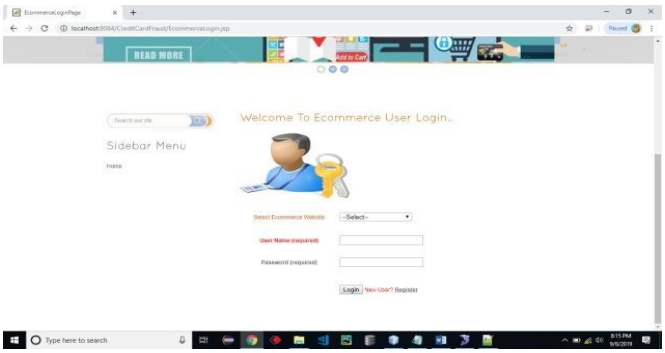


Fig.No.3 Login Page

It is the main login page of credit card fraud detection which defers with respect to the requirements of the user and admin. It also contain bank login page for both. admin and user.

XIII. CONCLUSION

Thus we have tried to implement the paper “Kuldeep Randhawa1, Chu Kiong Loo1”,”Creditcard fraud detection using Adaboost and Majority voting”IEEE 2016.As per the accomplishment system has checked whether the customer transactions are real and fake. Also this system adds to the authenticity and provides security to e-commerce.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

REFRENCES

[1].Maes S, Tuyls K, Vanschoenwinkel B, Manderick B. Master card fraud detection using Bayesian and neural networks. In Proceedings of the first international naiso congress on neuro fuzzy technologies 2006 Jan 16.

[2].Srivastava, Abhinav, et al. "Credit card fraud detection using hidden Markovmodel." IEEE Transactions on dependable and secure computing 5.1 (2008).

[3].Randhawa, Kuldeep, Chu Kiong Loo, ManjeevanSeera, CheePeng Lim, and Asoke K. Nandi. "Credit card fraud detection using AdaBoost and majority voting." IEEE access 6 (2018)

[4].Pumsirirat A, Yan L. Master card fraud detection using deep learning supported auto-encoder and restricted Boltzmann machine applications (2018).

[5]. Bhattacharyya, Siddhartha, et al. "Random Forest for Master card fraud: A comparative study." Decision Support Systems 50.3 (2011).