**GRAPHIC ERA HILL UNIVERSITY, DEHRADUN**

A picture containing text, room, gambling house

Description automatically generatedA picture containing text, clipart

Description automatically generated

**REPORT ON CREDIT CARD FRAUD DETECTION**

**2022-23**

**SUBMITTED TO: SUBMITTED BY:**

**Dr. Indrajit Kumar ABHYUDAY PRATAP SINGH**

**(Assistant Prof.) Semester: V**

**U. Roll No.: 2118034**

**Section: M**

***CREDIT CARD FRAUD DETECTION***

**CANDIDATE DECLARATION**

I hereby declare that the work which is presented in this project report on “Credit Card Fraud Detection” is done by me in the partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering of the Graphic Era Hill University.

Name: Abhyuday Pratap Singh

Roll no.: 2118034

**CONTENT**

1. ABSTRACT
2. INTRODUCTION
3. METHODOLOGY
4. EVALUATION AND ANALYSIS OF RESULT
5. CONCLUSION
6. REFERENCES
7. **ABSTRACT**

Detection of the Fraud of Credit Card is one of the maximum popular troubles in cutting-edge worldwide. that is due to the upward push of buying and selling and e-exchange structures. credit score card fraud normally takes location at the same time as a card is stolen for fraudulent capabilities or on the equal time as a fraudster makes use of credit score card information for their very own purposes. In worldwide, we are going through many credit score card issues. A credit score card fraud detection device is installed area to come upon fraudulent sports activities. This venture is broadly talking supposed to interest on machine reading algos. The algos used are RF set of on-line and Ada-boost set of online. The consequences of the 2 algos are primarily based on the accuracy, the precision, the maintain in mind, and the F1 rating. The ROC is drawn primarily based completely on the CM. We evaluate the Random wooded place set of online tips and the Ada-boost set of regulations and recall the set of on lines with awesome accuracy, precision, consider, and the F1 rating because the extraordinary set of on line tips used for fraud detection.

1. **INTRODUCTION**

In present day international, over-reliance at the internet is chargeable for the boom in credit card fraud transactions, however fraud isn't always only on-line, but also offline. Transactions also are increasing. records mining techniques [6] are used, but the effects aren't very accurate in detecting those credit score card fraud. The way best to limit these losses is to come across fraud with the efficient algorithms. that is a promising way to reduce the fraud of CC. With the spread of the internet, credit score playing cards are issued by way of financial corporations. you may borrow cash with a credit card. budget can be used for any reason. A condition while the cardboard is issued is that the cardholder repay the amount at the start borrowed and any agreed surcharges. Your credit score card is considered fraudulent if a person uses it for your behalf without your authorization. Scammers steal credit card PINs and account information and carry out fraudulent transactions without stealing the authentic physical card. credit score card fraud detection allowed us to decide whether new transactions were fraudulent or true.

Fraud can influence cards along with credit score and debit playing cards. the cardboard itself acts as a fraudulent supply in transactions. The cause of a criminal offense is to receive goods without deciding to buy them or to acquire cash fraudulently. credit score playing cards are clean goals for the fraud. The reason that you could make number of money in a completely quick time without taking a whole lot of dangers and it can even take weeks to detect against the law

Fig.1 Growth of e-commerce websites [12]

rapid growth of net usage [Figure 1] opportunities for fraudsters to dedicate credit card fraud are many The main scam taking place within the international today is e-trade internet site fraud. In present day technology, people are more inquisitive about getting things online than buying them. because of this, the increase of e-trade websites is accelerating and there's a large capability for credit card fraud. consequently, to keep away from such CCF, we need to find first-rate algorithms to reduce CCF.

1. **METODOLOGY**

The principal purpose of this report is to categorise transactions into information, which includes fraudulent and non-fraudulent transactions, the usage of algorithms such as the Random woodland and Ada-boost algorithms. We then examine these two algorithms and choose the first-rate algorithm for detecting fraudulent credit card transactions. The credit score fraud detection drift includes facts splitting, version schooling, model deployment, and scoring standards.

A detailed architectural diagram of a CCFD machine [Fig 2.] consists of many steps from collecting statistics units, developing models, and appearing analysis based totally at the effects. This model uses Kaggle's credit score card fraud facts. This dataset wishes to be pre-processed. To put together the model, we want to break up the records into education and look at facts. put together the Random woodland and Ada-boost fashions with education statistics. Then broaden each fashion. eventually, accuracy, precision, don't forget, and F1 ratings for each fashion are calculated. eventually, the assessment of credit score card fraud transactions could be more accurate.

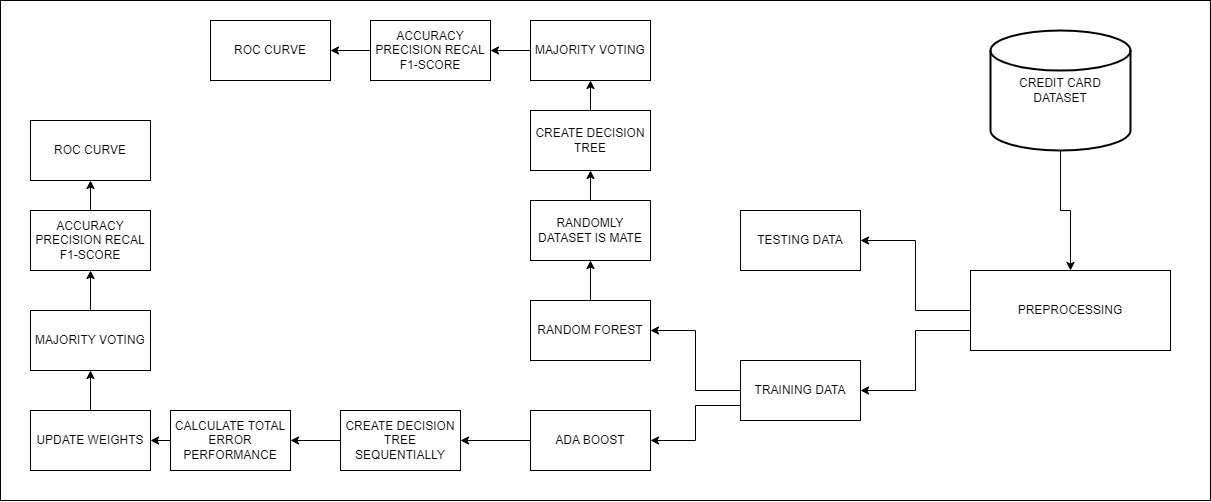


Fig.2 Architecture Diagram

1. The Algorithm of Random Forest

Random forest set of rules [Fig. 3] is the extensively used supervised learning (SL) algorithms. it could be used for each regression and type functions. however, the set of rules is particularly used for type issues. Forests are typically made of bushes. further, random forest algorithms assemble choice trees based on pattern facts and get predictions from all pattern statistics. 2d, the random forest area set of policies is an approach of ensemble. This set of policies is superior to single choice wooden because it reduces overfitting via averaging the results.

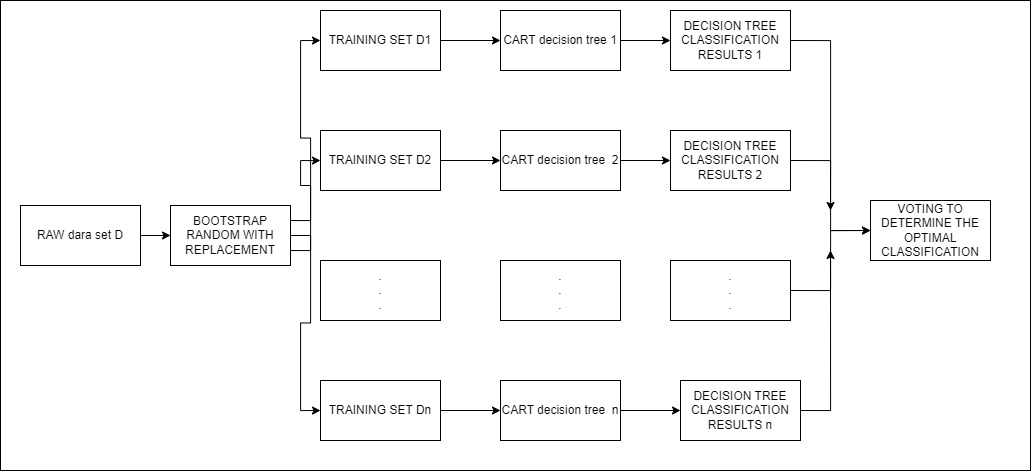


Fig 3 Random Forest Algorithm

Steps for Random Forest set of rules.

1. Take the educated Kaggle CC fraud statistics set and randomly select some sample facts.

2. the usage of randomly generated pattern statistics, construct a selection tree that is used to classify instances into fraudulent and non-fraudulent instances

3. A D- tree is shaped by splitting the nodes so that the node with the very best statistics gain becomes root node to classify the fraud and the non-fraud instances.

4. A majority vote is achieved right here, and decision tree may also return zero as an output. because of this those are non-fraudulent instances.

5. in the end, discover the accuracy, precision, don't forget, and the F1 score for both dishonest and non-cheating cases.

B. Ada-Boost Algo

Boosting is the technique of the ensemble. This algo is used to assemble robust classifier from the weak classifier. this can completed by means of using the weaker model of the series to create a more potent model. First, create a version from schooling records. His second model is then made out of the first model by way of fixing the errors that existed inside the previously created version. that is an iterative procedure that continues till both the most range of model are added, or the overall schooling set is predicted efficiently. Ada-boost is a hit boost algorithm developed for binary class.

Ada-boost stands for Adaptive Boosting. notable for difficult newbies. This Ada-boost boosting method [Fig. 4] combines several vulnerable classifiers into her one sturdy classifier. The Ada-boost set of rules can be used with short preference bushes. The manner AdaBoost is built is to first create the nodes, construct the tree, and then check the overall performance of the tree on each example. Weights also are given. schooling information that is hard to are looking ahead to is given better weight. The Ada-boost set of rules is a effective classifier that works nicely on both fundamental and complicated problem. The dis-advantage of this set of policies is that it's miles liable to noisy records. The set of policies is likewise touchy to outliers.

Steps Ada-boost Algo

1. The facts that Kaggle credit card fraud are acquired and skilled. Randomly pick out pattern records.

2. Use random sample records to sequentially construct decision timber to categorise fraud and the non-fraud instances.

3. A decision tree is shaped first. this will be performed by splitting the node based totally on the best records benefit, making it the foundation node, and classifying the fraudulent and non-fraudulent instances.

4. subsequent, calculate the error fee and the overall performance and replace the weights of incorrectly classified fraudulent and non-fraudulent transactions.

5. A majority vote is completed & the selection tree results in a result that indicates a non-fraudulent case.

6. The selection tree can output 1. This suggests a terrible case.

7. ultimately, discover the accuracy, precision, take into account, and F1 score for each cheat & non-dishonest cases.

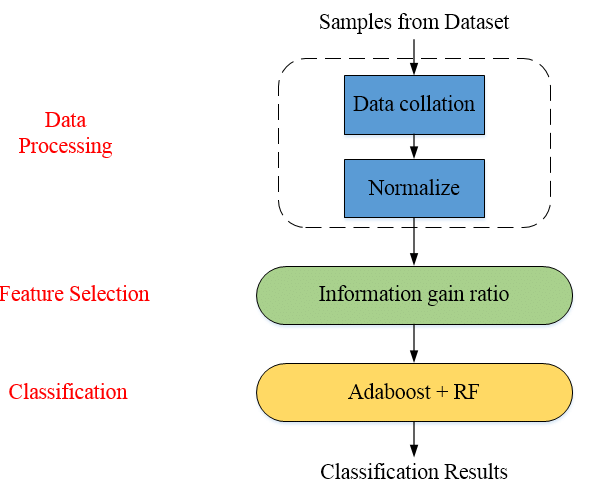
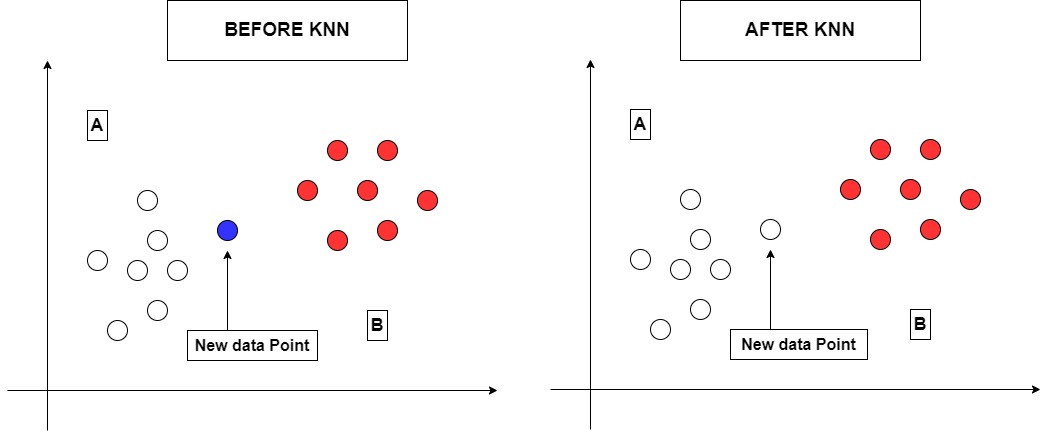


Fig 4 Ada-boost Algo

C. K-Nearest Neighbour (KNN) Algo

KNN come under a completely special sort of category of system learning algorithms which is probably known as ‘Lazy learners’ due to the truth this set of policies learns very slowly than in comparison to other algorithms. KNN follows a technique to research wherein it keeps specializing in storing the information until it's far having the input information whose label or elegance is meant to be predicted. KNN classifier predicts that how near the unidentified tuple is to the okay education set, and KNN does this through the use of a long way degree. In conceptual terms, KNN is straightforward and though capable of fixing complex problems as it may paintings fantastically with little information too.



1. **EVALUATION AND ANALYSIS OF RESULTS**

A. Dataset

credit score card fraud information are supplied through eu credit score card businesses. The dataset is pulled from Kaggle. This dataset includes transactions from credit score card holders. The dataset consists of transactions remodelled days. The dataset includes 284,807 transactions, of which 492 are fraud the ones fraud transaction account for only the 0.172% of all the transaction. Datasets with input variables are transformed to numeric values through PCA transformation. that is completed for confidentiality. The amount and time houses cannot be PCA transformed. The time beauty represents distinction in the elapsed second among every transaction and the primary. the amount elegance represents a economic transaction that has come about. a few other essential properties, "splendor", shows whether the transaction is fraudulent. pretty a variety of 1 indicates the fraud transaction and a 0 indicates a non-fraudulent transaction.

B. Criteria for Evaluation

To examine one-of-a-type algorithms, metrics together with accuracy, precision, consider, and F1 rating ought to be evaluated. A CM is also achieved. The CM is a matrix of 2\*2. The matrix consists of his 4 outputs TPR, TNR, FPR and FNR. Measures which encompass sensitivity, specificity, precision, and blunders fee can derived from the CM. Then discover the quality way to understand CCF.

The output of CM ---

1. authentic nice fee. it can be described because the quantity of fraudulent transactions taken into consideration fraudulent through the device.

2. genuine poor charge. it can be defined as the variety of valid transactions considered valid by means of the system.

3. fake wonderful price, which may be described as several the prison transactions which are not correctly categorized as a fraud.

4. fake bad price is described as a transaction incorrectly categorized as fraudulent however legitimate.

A receiver overall performance curve is built via manner of plotting TPR and FPR. this will be finished with specific thresholds. The ROC curve is a graph with TPR on the vertical axis and FPR on the horizontal axis.

C. Result

Confusion matrixes (CM) and ROC curves are plotted for the each. algorithm data go back special outputs when carried out to one-of-a-kind algorithms. here are the outcomes after first applying the random forest model to the dataset.

A screenshot of a computer

Description automatically generated

Fig.6 Output for Random Forest

I’m able to explain the scoring standards [Fig. 6]. Accuracy, reputation cost, and the F1 score are the equal for without fraud instances and extraordinary for cases of the fraud.

Graphical user interface

Description automatically generated

Fig 7 Confusion Matrix for RF

The mat [Fig 7] suggests the education facts consists of 190490 actual positives and zero false positives, and 0 proper negatives and 330 fake negatives.

Right here the Ada-boost set of rules statistics set is carried out. It offers effects like the random forest set of rules. See in fig[8]

A screenshot of a computer

Description automatically generated

Fig 8 shows the output from Ada-boost.

A graph of a number of different colored squares

Description automatically generated

Figure 9 Ada-boost Confusion Matrix

A diagram of a graph

Description automatically generated

Figure 10 KNN Confusion Matrix

Table

Description automatically generated

Fig 11 shows the output from KNN.

Here we display an evaluation between the random forest, Ada-boost and the KNN set of rules [Figure 1]. 12]. the two algorithms have the same accuracy, however, differ in accuracy, seek, and F1 rating. The random forest algorithm has the best accuracy, don't forget, and F1 rating.

Fig 12 Comparing Algorithms

1. **CONCLUSION**

There are several exclusive strategies for fraud detection, but no unmarried algorithm is great for fraud detection. From evaluation, we say that the accuracy is same for RF and the boost algo AND KNN rules set. The RF rules set outperforms the Ada-boost algorithm in terms of accuracy, recall, and the score of F1 but The KNN rules set outperforms the RF algorithm in terms of accuracy, recall, and the score of F1. From this, we finish that the KNN set of rules outperforms the Ada-boost set of rules in detecting CCFD.

1. **REFERENCES**
2. Adi Saputra1, Suharjito2L: Fraud Detection using Machine Learning in e-Commerce, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.
3. Dart Consulting,Growth Of Internet Users In India And Impact On Country’s Economy: https://www.dartconsulting.co.in/marketnews/growth-of-internet-users-in-india-and-impact-on-countryseconomy/
4. Ganga Rama Koteswara Rao and R.Satya Prasad, “ -Shielding The Networks Depending On Linux Servers Against Arp Spoofing, International Journal of Engineering and Technology(UAE),Vol. 7, PP.75-79, May 2018, ISSN No: 2227-524X, DOI - 10.14419/ijet.v7i2.32.13531.
5. Heta Naik, Prashasti Kanikar: Credit card Fraud Detection based on Machine Learning Algorithms,International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 44, March 2019.
6. Navanshu Khare, Saad Yunus Sait: Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models, International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 825-838 ISSN: 1314-3395.
7. Randula Koralage, , Faculty of Information Technology, University of Moratuwa,Data Mining Techniques for Credit Card Fraud Detection.
8. Roy, Abhimanyu, et al:Deep learning detecting fraud in credit card transactions, 2018 Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2018.
9. Sahayasakila.V, D. Kavya Monisha, Aishwarya, Sikhakolli VenkatavisalakshiseshsaiYasaswi: Credit Card Fraud Detection System using Smote Technique and Whale Optimization Algorithm, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-8 Issue-5, June 2019.
10. Statista.com. retail e-commerce revenue forecast from 2017 to 2023 (in billion U.S. dollars). Retrieved April 2020, from India:
11. Yashvi Jain, Namrata Tiwari, Shripriya Dubey, Sarika Jain: A Comparative Analysis of Various Credit Card Fraud Detection Techniques, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-5S2, January 2019.
12. Yong Fang1, Yunyun Zhang2 and Cheng Huang1, Credit Card Fraud Detection Based on Machine Learning, Computers, Materials & Continua CMC, vol.61, no.1, pp.185-195, 2019.
13. Kaithekuzhical Leena Kurien, Dr. Ajeet Chikkamannur: Detection And Prediction Of Credit Card Fraud Transactions Using Machine Learning, International Journal Of Engineering Sciences & Research Technolog.
14. Asha RB, Suresh Kumar KR: Credit card fraud detection using artificial neural network, Global Transitions Proceedings, 2021.
15. A.A. Taha, S.J. Malebary, An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine, IEEE Access 8 (2020) 25579–25587, doi:10.1109/ACCESS.2020.2971354.
16. Joy Iong-Zong Chen, Kong-Long Lai , Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert in 2021.
17. Fabrizio Carcilloa, Yann-Aël Le Borgnea , Olivier Caelenb , Yacine Kessaci , Frédéric Oblé , Gianluca Bontempi, Combining unsupervised and supervised learning in credit card fraud detection 2021.