International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Supervised Approach to Credit Card Fraud Detection Using Machine Learning

Reddi Sandeep Kumar

Student, Rajam, Vizianagaram, 532127, India.

ABSTRACT

Credit card fraud refers to the physical loss of master card or loss of sensitive master card information. Credit card fraud detection is presently the foremost frequently occurring problem within the present world. Now a days Credit card is the commonly used payment mode. Because the technology is developing, the amount of fraud cases are also increasing. But due to lot of loop holes in this system, many problems are arising in the method of credit card scams. Due to this the companies as well as customers who are using credit cards face an enormous loss. Individuals master card information is collected illegally and it is used for fraudulent transactions. To detect the fraudulent activities the master card fraud detection system was introduced. Generally we have many machine learning algorithms that are used for credit card fraud detection. Random Forest, Logistic Regression, Adaboost are a number of the machine learning algorithms used for Credit Card Fraud Detection. Using these algorithms we calculate the accuracy, precision, recall. The algorithm that has the best accuracy, precision and recall is taken in to account as the best algorithm that is used to detect the fraud.

Keywords**:** Credit Card Fraud, Random Forest, Logistic Regression, Machine Learning, Adaboost, Accuracy, Precision.

1. Introduction

Finding fraudulent credit card transactions is our main goal. It is necessary to categorise the fraudulent and non-fraudulent transactions in order to do this. The main objective is to create a fraud detection programme that uses machine learning-based classification methods to quickly and accurately identify transactions that are fraudulent. As technology develops quickly, less money is paid in cash and more money is sent online, which makes it easier for fraudsters to conduct anonymous transactions. Only the card number, expiration date, and CVV are needed in some online payment methods and in some situations, without even our knowledge, the data is being stolen. When someone else uses your credit card without your permission in your place, it is referred to as credit card fraud.

Without taking the original physical card, fraudsters can carry out any illicit transactions by stealing the PIN or account information from the credit card. We could determine whether the new transactions are genuine or fraudulent using the credit card fraud detection. The best technique to determine if a transaction is fraudulent or not is to determine the customer's spending habits using available data and machine learning algorithms to determine whether the transaction is real or not. Billion-dollar financial losses result from the use of credit cards without adequate security. Global financial losses as a result of credit card theft total 22.8 billion US dollars in 2020, and by 2022, the number is anticipated to rise steadily to 31 billion US dollars.

2. Literature Survey

In paper [1] Bhanusri, Andhavarapu, K. Ratna Sree Valli discuss about various credit card fraud detection models. In addition to being the most widely used method of payment for both normal and online purchases, credit card theft is also on the rise. Fraud is any malicious action intended to harm the other party financially. We need a strong fraud detection system that not only finds the fraud, but also finds it accurately and before it happens in order to stop it. Additionally, we must ensure that our systems are capable of learning from previous frauds that have been perpetrated and adapting to new fraud techniques in the future. The idea of credit card fraud and its different variants have been introduced in this paper. The Support Vector Machine (SVM), Artificial Neural Networks (ANN), Bayesian Network, Random Forest(RF), Adaboost, and Decision Trees are just a few of the methodologies we've covered for fraud detection systems.

In paper [2] . A. A. Taha and S. J. Malebary, discuss the amount of pertinent research efforts in the literature has risen as a result of the potential social and economic significance of identifying fraudulent credit card transactions. This study suggests a method for using an enhanced light gradient boosting machine to detect fraud in credit card transactions (OLightGBM). Historical credit card transactions are classified as valid or fraudulent using supervised learning algorithms. Also, supervised literacy algorithms start learning using these data to produce a model that can be used to classify new data samples.Bayesian belief networks (BNNs) and decision trees (DTs) were used to detect fraud in financial transactions. Here, a data set of financial transactions collected from 76 Greek industrial companies was used. The BNNs obtained the highest accuracy (90.3%), whereas the DTs achieved an accuracy of 73.6% .

In paper [3] D. Varmedja, M. Karanovic, S. Sladojevic, Researchers were driven to develop a method to identify and stop frauds by the significant loss that fraudulent operations are creating. Many approaches have already been developed and examined. Below is a brief review of some of them. Traditional methods have shown to be effective, including Gradient Boosting (GB), Support Vector Machines (SVM), Decision Tree (DT), LR, and RF. On a European dataset, this study's employment of GB, LR, RD, SVM, and a combination of specific classifiers produced high recall of over 91 percent. Only after balancing the dataset by undersampling the data were high precision and recall attained.In this study, the models based on LR, DT, and RF were compared while also using a European dataset. The best of the three models, RF, had accuracy of 95.5 percent, followed by DT's accuracy of 94.3 percent and LR's accuracy of 90 percent.

In paper [4]]. Sharma, Pratyush, Souradeep Banerjee, Devyanshi Tiwari discussed there are many techniques available to detect fraud transactions. It is very difficult to detect the fraud, or they can be detected after the fraud happens. This happens because the fraudulent transactions are small as compared to total transactions. Random forest is a decision tree regression and classification technique that works well with both categorical and numerical data. The authors tested random forest and SVM classifier to detect fraudulent transactions from the dataset. The pre-processing was done to avoid missing values and scale feature values. The authors concluded that imbalanced data did not work well with SVM as compared to random forest classifiers. Advantage of using the random forest technique was the introduction of new data points did not have a major impact on the model since it used a subset of data with different decision trees. Authors compared colorful styles like decision trees and arbitrary timber and set up that arbitrary timber classifier proves better than decision trees and logistic retrogression for the delicacy for logistic retrogression, Decision tree, and arbitrary timber classifier are90.0,94.3, and95.5 independently.

In paper [5] E. Ileberi, Y. Sun and Z. Wang, discussed In this paper, we implement machine learning (ML) algorithms for credit card fraud detection that are evaluated on a real world dataset which was generated from European cardholders in September 2013. This dataset is highly imbalanced. To alleviate the issue of class imbalance that is found in the European card dataset, this research investigated the use of the Synthetic Minority Over-sampling Technique (SMOTE). Moreover, the ML methods that were considered in this research include: Support Vector Machine (SVM), Random Forest (RF), Extra Tree (ET), Extreme Gradient Boosting (XGBoost), Logistic Regression (LR), and Decision Tree (DT). These ML methods were evaluated individually in terms of their effectiveness and classification quality. Additionally, the Adaptive Boosting (AdaBoost) algorithm was paired with each methods to increase their robustness. The main contribution of this paper is a comparative analysis of several ML methods on a publicly available dataset that contains real word cards transactions. The results demonstrated that the AdaBoost-SVM achieved an accuracy of 99.959% .

3. Data Collection

The below data is collected from the references.

4. Methodology

Here we use three different machine learning algorithms. They are Random Forest Algorithm, Naïve Bayes and Decision Tree algorithms. Then we compare the three algorithms performances using confusion matrix and ROC curves. Which algorithm will give the best accuracy and precision will take into the consideration.

Steps for Random Forest Algorithm:

Pick some sample data at random from the trained Kaggle credit card fraud dataset. The Decision Trees that are used to categorise the cases into fraud and non-fraud cases are now formed using the randomly generated sample data.

The root node of the Decision Trees, which classify fraud and non-fraud situations, is generated by separating the nodes; the nodes with the biggest Information Gain become these nodes.

Now that the majority vote has been conducted, the decision trees may produce an output of 0, indicating that these are circumstances where fraud has not occurred. The accuracy, precision, and recall are finally determined for both fraud and non-fraud cases.

METHODOLOGY:

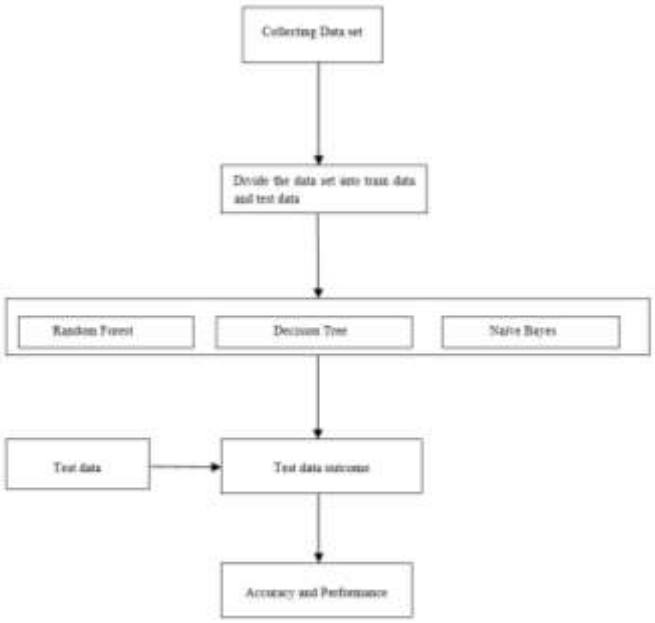

Fig: Model architecture

Step 1: Import raw data set

Step 2: Convert raw data into data frames .

Step 3: Perform a random sample

Step 4: Decide how much data will be used for training and testing.

Step 5: Give 70% of the data for training and the remaining 30% for testing.

Step 6: Give the models the training dataset.

Step 7: Use the algorithm to build the model while applying it to three other algorithms.

Step 8: Making predictions for the test dataset for each method

Step 9: Determine each algorithm's accuracy.

5. Results and Discussion

Random Forest A large number of decision trees are erected during the training phase of the arbitrary timbers or arbitrary decision timbers ensemble literacy approach, which is used for bracket, retrogression, and other tasks. The class that the maturity of the trees chose is the affair of the arbitrary timber for bracket problems. The mean or average vaticination of each individual tree is returned for retrogression tasks. The tendency of decision trees to overfit their training set is corrected by arbitrary decision timbers. Although they constantly outperform decision trees, grade boosted trees are more accurate than arbitrary timbers. still, their effectiveness may be impacted by data tricks.

Naive Bayes The Naive Bayes algorithm is a supervised literacy system for bracket issues that's grounded on the Bayes theorem. It's substantially employed in textbook categorization with a large training set. One of the most straightforward and effective bracket algorithms is the Naive Bayes Classifier, which aids in the development of quick machine literacy models able of making accurate prognostications. Being a probabilistic classifier, it makes prognostications grounded on the liability that an object will do. Spam filtration, novelettish analysis, and categorising papers are a many common operations of the Naive Bayes algorithm.

Decision Tree A supervised literacy system called a decision tree can be used to break bracket and retrogression problems, but it's generally favoured for doing so. It's a tree- structured classifier, where internal bumps stand in for a dataset's features, branches for the decision- making process, and each splint knot for the bracket result. The Decision knot and Leaf Node are the two bumps of a decision tree. While Leaf bumps are the results of opinions and don't have any further branches, Decision bumps are used to produce opinions and have multitudinous branches. The given dataset's features are used to execute the test or make the opinions.

The suggested system's performance is assessed using the F1 score, precision, recall, and accuracy. The accuracy of the system is 0.9793, as shown in the suggested system output in Figure. This demonstrates that the suggested technique had displayed greater accuracy for a significant amount of traing data.

```
Confusion matrix :
 [[19451    34]
 [  379   136]]
Outcome values :
 tp= 136 fn= 379 fp= 34 tn= 19451
Classification report :
              precision    recall  f1-score   support

           1       0.80      0.26      0.40       515
           0       0.98      1.00      0.99     19485

    accuracy                           0.98     20000
   macro avg       0.89      0.63      0.69     20000
weighted avg       0.98      0.98      0.97     20000

rf acc is  0.97935
```

Fig: Performance Evaluation of Proposed System

Comparison of Performance Figure illustrates how the proposed system compares to two other classifiers, Decision Tree and Naive Bayes, in terms of performance. It is obvious that our suggested solution, which used the Random Forest technique, outperformed Decision Tree andNaveBayesTechnique.

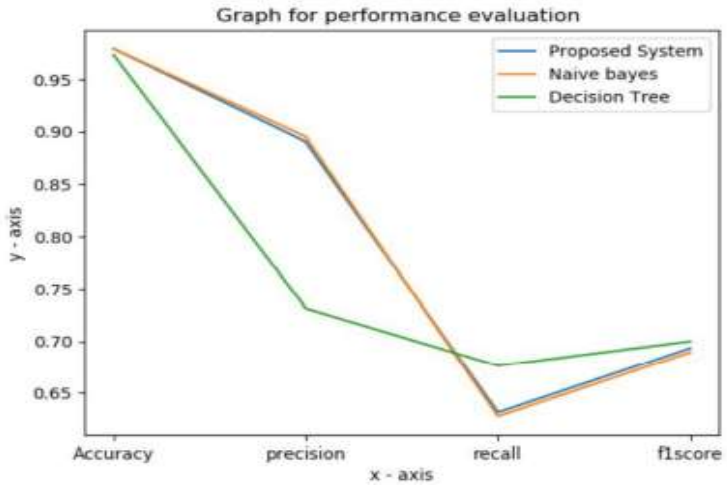

Fig: Comparative Performance Evaluation of Proposed System

6. Conclusion

Credit card frauds represent a veritably serious business problem. These frauds can lead to huge losses, both business and particular. Because of that, companies invest further and further plutocrat in developing new ideas and ways that will help to descry and help frauds. The main thing of this paper was to compare certain machine learning algorithms for discovery of fraudulent deals and also to address the different machine learning algorithms and how they can be employed in different ways to descry fraud. In the future, we can ameliorate our classifier so that can get close to the thing of 100 delicacy. Multiple algorithms can be composite together, and their results can be compounded to ameliorate the overall delicacy of the system. In this paper, we studied operations of machine literacy like Naïve Bayes, Decision trees, Random timber shows that it proves accurate in abating fraudulent sale and minimizing the number of falsealerts.However, the probability of fraud deals can be prognosticated soon after credit card deals, If these algorithms are applied into bank credit card fraud discovery system. And a series ofanti-fraud strategies can be espoused to help banks from great losses and reduce pitfalls. - score, support and delicacy are used to estimate the performance for the proposed system. By comparing all the three styles, we set up that arbitrary timber classifier fashion is better than the decision tree and naïve bayes styles.

REFERENCES

[1]. Bhanusri, Andhavarapu, K. Ratna Sree Valli, P. Jyothi, G. Varun Sai, and R. Rohith. "Credit card fraud detection using Machine learning algorithms." Journal of Research in Humanities and Social Science 8, no. 2 (2020): 04-11.

[2]. A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," in IEEE Access, vol. 8, pp. 25579-25587, 2020, doi: 10.1109/ACCESS.2020.2971354.

[3]. D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. Anderla, "Credit Card Fraud Detection - Machine Learning methods," 2019 18th International Symposium INFOTEH-JAHORINA(INFOTEH), 2019, pp. 1-5, doi: 10.1109/INFOTEH.2019.8717766.

[4]. Sharma, Pratyush, Souradeep Banerjee, Devyanshi Tiwari, and Jagdish Chandra Patni. "Machine learning model for credit card fraud detection-a comparative analysis." Int. Arab J. Inf. Technol. 18, no. 6 (2021): 789-796.

[5]. E. Ileberi, Y. Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," in IEEE Access, vol. 9, pp. 165286-165294, 2021, doi: 10.1109/ACCESS.2021.3134330.

[6]. J. Gao, B. Ding, W. Fan, J. Han, P.S. Yu, ―Classifying data streams with skewed class distributions and concept drifts‖, IEEE internet comput., vol.12, no. 6, pp. 37-49, Nov 2008

[7]. E. Aleskerov, B. Freisleben, and B. Rao, ―CARDWATCH: A neural network based database mining system for credit card fraud detection,‖ in Proc. IEEE/IAFE Computat. Intell. Financial Eng., Mar. 1997, pp. 220–226.

[8]. Yashvi Jain, Namrata Tiwari, ShripriyaDubey, Sarika Jain, A Comparative Analysis of Various Credit Card Fraud Detection Techniques, Blue Eyes Intelligence Engineering And Sciences Publications 2019

[9]. Yashvi Jain, NamrataTiwari, ShripriyaDubey,Sarika Jain:A Comparat ive Analysis of Various Credit Card Fraud Detect ion Techniques, Internat ional Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-5S2, January 2019.

[10]. Roy, Abhimanyu, et al:Deep learning detecting fraud in credit card transactions, 2018 Systems and Informat ion Engineering Design Symposium (SIEDS), IEEE, 2018.