JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Credit Card Fraud Detection Using Machine Learning Techniques

Prof. Sheetal Nirve
Department of Computer Engineering
K J College of Engineering and Management Research, Pune
Pune, India

Prajakta More
Department of Computer Engineering Engineering
K J College of Engineering and Management Research, Pune
Pune, India

Pratiksha Kanje
Department of Computer
K J College of Engineering and Management Research, Pune
Pune, India

Riya Wagh
Department of Computer Engineering
K J College of Engineering and Management Research, Pune
Pune, India

Abhijeet Chakane
Department of Computer Engineering
K J College of Engineering and Management Research, Pune
Pune, India

Abstract: **The widespread use of credit cards has led to more cases of fraud. Credit cards have made online shopping and electronic payments easier, but they've also made it easier for fraudsters to steal money. To combat this, we're using computer programs called machine learning to find and stop fraud. These programs are good at analyzing information about customers. Credit card fraud has been increasing in recent years, causing financial problems for cardholders, merchants, and banks. This review paper looks at different ways to detect fraud using machine learning and compares them by how well they work. The paper suggests a new system that uses a method called supervised Random Forest to improve the accuracy of detecting credit card fraud.**

Keywords: *Credit card frauds, Machine Learning, Random Forest Algorithm, Artificial Neural Network(ANN)*

1. INTRODUCTION

Credit cards are a popular method of payment for Internet purchases because they're simple and practical. However, there is a rise in credit card fraud and abuse as their use increases. Problems arise from credit card fraud for both the cardholders and the issuers of the cards.

Fraudsters, or the people who commit fraud, are getting smarter and finding new ways to steal money without getting caught. Credit card fraud can happen when someone uses a credit card without permission, does weird or unusual transactions, or uses a card that's not active anymore. Credit card fraud has been happening more and more in recent years, especially with the rise of online shopping and electronic banking.

In this research study, we want to figure out how to stop these frauds. We're looking at things like public information, data imbalances (which means there's a lot more of one type of data than another), changes in how fraud happens, and too many false alarms (when the system thinks there's fraud but there isn't). Our goal is to find better ways to detect and prevent credit card fraud in this digital age.

Machine Learning is a really good way to catch fraud. It uses two main methods: one where it learns from examples with labels called supervised learning (like teaching it what is definitely fraud and what's not), and another where it groups customers based on how they usually spend money to find unusual behavior that might be fraud called unsupervised learning. So, it's like teaching a computer to spot suspicious activity on credit cards. There are different types of fraud which are as follows:

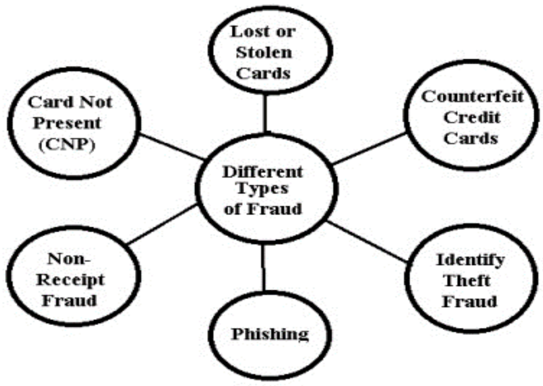

Figure 1: Different Types of Fraud

Many different computer methods have been used to find credit card fraud, like neural networks, decision trees, and more. This paper looks at how well these methods work and compares them. They evaluate each other's performance in terms of determining if a transaction is legitimate or fraudulent. The outcome demonstrates that, when compared to the other algorithms, the Random Forest Algorithm is the most exact and accurate at identifying fraud.

2. RELATED WORK

Several algorithms exist; a few of them are discussed below: Through the use of three distinct approaches—CNN, LSTM, and Auto-encoders (AEs)—the research seeks to enhance the detection of credit card fraud. They put four distinct models to the test: AE, CNN, LSTM, and AE&LSTM.With an excellent accuracy score of 0.98, the AE model was the most accurate. Both the CNN and LSTM models had a decent accuracy of 0.86. After 400 attempts, the AE&LSTM model had the lowest accuracy, at just 0.30[1]. The authors discuss the issues with the state-of-the-art fraud detection systems and describe how they developed a more intelligent system based on the ABC optimization method. They want to fix the issues with fraud detection. Instead of using simple methods like data mining, they've created a system that combines different techniques to make it work better. They use a rule engine to pick out important information from a large set of data. The goal is to be more accurate and save time and money when detecting fraud. They also plan to improve the system in the future by adding more rules to make it even better at catching fraud.[3] So In the study of using a dataset from Kaggle, researchers found that the new methods are good at finding frauds without making too many mistakes. They made the data better by preprocessing it (cleaning it up), and this made the methods work even better. Out of all the methods they tried, K-Nearest Neighbor (KNN) was the best at finding frauds. They measured it by looking at how accurate it was, how many frauds it found, and how many it missed.[8] When authors talk about prediction, In data mining, there are two main jobs: figuring out what's important in the data (feature extraction) and deciding what category something belongs to (classification). For catching credit card fraud, researchers have come up with different ways to do this classification. But, sometimes the methods used to pick out important things from the data aren't very good at showing how things are connected. This can make the classification part not work so well. In the paper, they used two methods, K-Nearest Neighbor (KNN) and Naïve Bayes, to pick out important stuff from the data and then decide if something is fraud or not. They did this using a computer program written in Python. And their new approach improved fraud detection by about 99%. So, they discovered an improved method for identifying credit card fraud.[9] To identify fraudulent transactions, a variety of supervised machine learning techniques are applied including Decision Trees, Logistic Regression [LR], Naive Bayes Classification [NBC], and SVM. The paper presents a revolutionary fraud detection approach. In the model the customers are grouped based on their transactions and the cardholder's profile is built on their behaviour patterns. Different classifiers are applied to different types of groups of customers and then a rating score is generated for each type of classifier. Since oversampling yields inadequate outcomes, the SMOTE procedure is applied to the imbalance dataset. The cardholder's most recent behavioral pattern is determined by the classifier with the highest rating score.[11]. Three different machine learning methods were assessed by the researchers: K-Nearest Neighbors (KNN), Naïve Bayes, and Logistic Regression. They set out to evaluate these algorithms' effectiveness in spotting fraudulent credit card transactions as their main goal. The research was designed to ascertain which among these algorithms exhibited superior performance in the field of identifying frauds. To achieve this objective, investigators likely conducted a comprehensive analysis that encompassed the measurement and comparative assessment of various performance metrics, including but not limited to accuracy, precision, and other relevant indicators. The outcomes derived from that study bear significant implications for financial institutions and security professionals, offering valuable insights for the strategic selection of the most appropriate machine learning algorithm to bolster their credit card fraud detection systems. These findings can be instrumental when contextualized within the broader landscape of related research.[4] The study explores the use of ML techniques to forecast credit card fraud incidents in this dataset. Specifically, two distinct machine learning algorithms, namely the Decision Tree and Random Forest, were employed to analyze the dataset for fraud detection purposes. Notably, the Random Forest algorithm exhibited superior performance in comparison to the Decision Tree. However, a critical issue identified in the dataset was the significant class imbalance, where fraudulent cases were notably less prevalent than legitimate ones. In light of this imbalance, the Decision Tree algorithm was favored for implementation. Furthermore, the research took measures to address the class imbalance problem, opting for oversampling techniques. This strategy was employed to mitigate the imbalance issue and enhance the performance metrics, particularly accuracy scores. These observations and methodologies provide valuable insights into the credit card fraud detection domain, especially within the context of addressing class imbalance concerns, which can be of interest within the related work of similar studies.[10] The dataset underwent a data-splitting process, partitioning it into training data (70%) & testing data (30%). Subsequently, The dataset was analyzed and predicted using three different machine learning algorithms: Random Forest, Logistic Regression, and Naive Bayes. The Naive Bayes algorithm stands out for its efficiency in training and scalability, making it a notable choice for this study. Additionally, the AdaBoost algorithm, primarily designed for binary classification, was considered. It involves assigning different weights to each instance within the training dataset to enhance the model's performance.The research yielded promising results in the realm of accurately identifying fraudulent transactions while simultaneously minimizing false alarms. Moreover, the study demonstrated the ability to predict the likelihood of fraudulent transactions shortly after a credit card transaction takes place.Performance evaluation of the several criteria including accuracy, precision, recall and accuracy were used to evaluate the suggested system[12]. The research likely involves training of an artificial neural network using historical data related to credit card transactions, encompassing both legitimate and fraudulent instances. This training process enables the neural network to recognize patterns and characteristics associated with both types of transactions. The backpropagation technique is subsequently employed to iteratively adjust and optimize the neural network's internal parameters, thereby improving its ability to classify transactions accurately.

That aligns with ongoing endeavors aimed at bolstering the security of credit card transactions, ultimately safeguarding both individual cardholders and financial institutions against the perils of fraudulent activities. The integration of artificial neural networks and backpropagation represents a promising avenue in the pursuit of more effective credit card fraud detection strategies.[5] Neural networks represent computational systems modeled after the human brain's capacity to acquire knowledge and make decisions grounded in data analysis. The likely methodology employed entails the training of a neural network using historical datasets encompassing credit card transaction records. Through this training process, the neural network is instructed to discern and internalize patterns characteristic of both legitimate and fraudulent transactions. The primary objective is to establish an automated system capable of identifying suspicious transactions autonomously, thereby mitigating the susceptibility to fraudulent activities.[6] In the study, it is apparent that the authors delve into the utilization of the Random Forest algorithm, a prominent machine learning technique. The primary goal is to increase the identification of fraudulent CCF with greater accuracy and efficiency. The Random Forest algorithm is well-regarded for its proficiency in managing intricate data sets and delivering precise predictions. The research presumably entails the training of a Random Forest model using historical credit card transaction data. The model assimilates and comprehends the intricate patterns inherent in legitimate and fraudulent transactions. The overarching aim is to construct a resilient system endowed with the capability to autonomously discern potentially fraudulent activities, thereby bolstering the overall security posture of credit card transactions.[7] There are different supervised and unsupervised learning algorithms each with its unique characteristics and capabilities: Logistic Regression: This algorithm utilizes regression analysis to estimate parameters related to input data. The final output is determined by the logistic function curve, which is particularly useful for binary classification tasks. Support Vector Machine (SVM): SVM operates by plotting data points in a high-dimensional space (often 20 dimensions in this study) to maximize the margin distance between different classes. Its primary objective is to achieve effective separation between classes. Random Forest Classifier: This algorithm is versatile and applicable to both classification and regression problems. It functions by aggregating multiple decision trees to derive a final classification result, making it robust and resilient. Artificial Neural Network (ANN): ANN simulates the decision-making capabilities of neurons. It processes input data through a complex network of interconnected nodes, ultimately producing a single output per node. ANNs are renowned for their capacity to identify complicated trends in data.

The study's findings indicate that ANN performed exceptionally well, achieving a precision rate of 99.68% for fraudulent transaction detection. In contrast, SVM exhibited a relatively high false alarm rate of 5.2%, while the performance of the decision tree algorithm was average. Furthermore, the Random Forest algorithm demonstrated an improvement over the decision tree method[13]

3. COMPARISON

Based on all research compared all the algorithms which give results in the following table:

Algorithm	Accuracy	Precision	Recall
Decision Tree	0.86	0.84	0.88
Logistic Regression	0.90	0.88	0.91
Naive Bayes Classifier	0.84	0.82	0.85
SVM	0.92	0.90	0.93
K-Nearest Neighbors	0.88	0.86	0.89
Random Forest	0.96	0.92	0.95

Table 1: Comparison of algorithms

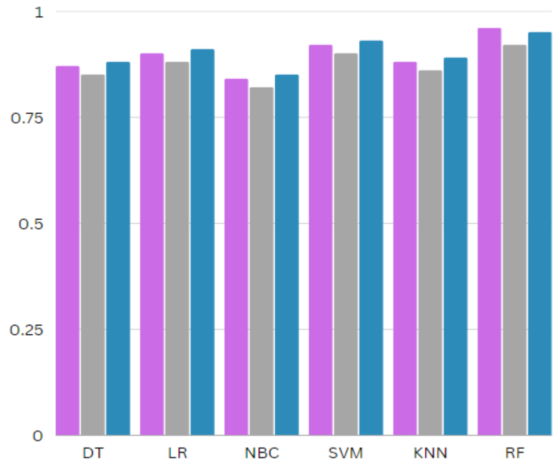

Figure 2: Performance of all algorithms

4. PROPOSED SYSTEM

Nowadays, credit cards are used for a variety of purposes. Unfortunately, though, credit card transaction fraud is becoming more and more common. An enormous amount of money is lost as a result of this fraud every year. Fraud can happen in different ways, and it keeps changing because criminals use new tricks with technology.

To fix this problem, we're going to use machine learning techniques that give clearer results in terms of accuracy. This way, we can stop fraudulent transactions and keep the money safe.

4.1 Objectives

The main objectives are as follows:

- To minimize financial losses for both the credit card company and its customers.
- To make the system user-friendly.

- Make a system that will be easy to maintain.
- To Reduce Credit Card Fraud.
- Faster detection of fraud.

5. CONCLUSION

In our research, we explored different computer methods to catch fraudulent credit card transactions. We checked how good they were at their job using measures like accuracy, precision, and recall. After going through all the options, we chose one called "Random Forest." This method is like having a smart helper that can tell if a credit card transaction seems fishy or not. It's a way of making sure that the money stays safe by catching suspicious activities with a clever computer system.

REFERENCES

[1]Arjwan H. Almuteer1 , Asma A. Aloufi1 , Wurud O. Alrashidi , Jowharah F. Alshobaili1 , Dina M. Ibrahim "Detecting Credit Card Fraud using Machine Learning" International Journal of Interactive Mobile Technologies (iJIM) – eISSN: 1865-7923 – Vol. 15, No. 24, 2021 https://doi.org/10.3991/ijim.v15i24.27355

[3] Darwish SM. An intelligent credit card fraud detection approach based on semantic fusion of two classifiers. Soft Comput. 2019;24:1243–53. https://doi.org/10.1007/s00500-019-03958-9.

[4] Itoo F, Meenakshi and SS. "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection". Int J Inf Technol. 2020;13:1503–11.

https://doi.org/10.1007/s41870-020-00430-y

[5] Dubey SC, Mundhe KS, Kadam AA. Credit card fraud detection using artificial neural network and backpropagation. In: 2020 4th international conference on intelligent computing and control systems (ICICCS). IEEE; 2020. p. 268–273.

[6] Patidar R, Sharma L. Credit card fraud detection using neural network. Int J Soft Comput Eng (IJSCE), 2011;1(32–38).

[7] Jemima Jebaseeli T, Venkatesan R, Ramalakshmi K. Fraud detection for credit card transactions using random forest algorithm. Singapore: Springer; 2020.

[8] Rucha Narkhede, Nilesh Chaudhari "Detecting Frauds In Credit Card Using KNN And Random Forest Machine Learning Approach"2022 IJCRT |ISSN: 2320-2882

[9] Shishobitveer Singh, Vinay Chopra "HYBRID MACHINE LEARNING ALGORITHM FOR CREDIT CARD FRAUD DETECTION" irjmets/ Volume:04/Issue:09/September-2022

[10] BORA MEHAR SRI SATYA TEJA1, BOOMIREDDY MUNENDRA2, Mr. S. GOKULKRISHNAN "A Research Paper on Credit Card Fraud Detection" irjet/ Mar 2022 e-ISSN: p-ISSN: 2395-0072

[11] Vaishnavi N D, Geetha S "Credit Card Fraud Detection using machine learning algorithms" 2019 by Elsevier B.V.

[12] K.Ratna Sree Valli, P.Jyothi, G.Varun Sai, R.Rohith Sai Subash "Credit Card Fraud Detection using machine learning algorithms" *Volume 8 ~ Issue 2 (2020)pp.: 04-11 ISSN(Online):2321-9467* www.questjournals.org

[13] Pratyush Sharma, Souradeep Banerjee, Devyanshi Tiwari, and Jagdish Chandra Patni "Machine Learning Model for Credit Card Fraud Detection- A Comparative Analysis" 2021/ https://doi.org/10.34028/iajit/18/6/6