WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Fraud Detection Using Data Mining Techniques

Shaik Ghouse Basha

Student, Rajam, Vizianagaram, 532127, India.

ABSTRACT

Any conduct that uses deception to obtain an advantage is referred to as "fraud." A crime is regarded to have been committed when a material fact was concealed, or a knowing lie was told. In other terms, fraud is when a person or organisation is cheated out of their money or property by lying. Nowadays, as technology develops, many people are deceived about it since they don't comprehend it. Frauds are growing increasingly common in today's culture, especially those involving credit and debit cards. The volume of consumer and bank frauds is astounding. because the general public lacks technological knowledge. Considering how frequently fraud occurs, this appears to be a major issue. There may be a number of various tactics employed to reduce fraud. In this, our primary focus is on developing the best algorithm that aids in preventing fraud, lowering the rate of fraud, or reducing miscommunication between banks and the general public that is likely to be a good indicator of the scope of fraudulent operations in the financial sector.

1. Introduction

A credit card is a large, practical plastic card that carries personal information, such as a signature or photo, card numbers, or magnetic stripe/chip data. It enables the person listed on it to make purchases or handle accounts in his name, for which he will occasionally be charged. One alternative to cash payment is a credit card. Some cardholders might misuse their ability to use and repay credit cards. In addition, credit card transactions are vulnerable to fraud, when unauthorised parties use credit cards to carry out illicit activities. Therefore, it is the obligation of card issuers or the banks to come up with a practical solution to lower any costs that may arise from the aforementioned problems. Data mining is one approach to solving these problems. They have a special CCV number that is quite important. The physical security of the plastic card and the safeguarding of the credit card number are both factors in its security [1]. The number of credit card exchanges is increasing quickly, which has led to a significant rise in fraudulent activity. The term "credit card fraud" refers to a broad range of theft and misrepresentation that uses a credit card as a fictitious source of funds in a particular transaction. To address this fraud recognition issue, measurable procedures and numerous information mining calculations are used in significant numbers. The bulk of artificial intelligence-based systems for credit card extortion rely on artificial intelligence, meta learning, and pattern matching. The use of any depiction's instalment card as a means of misrepresentation is just the beginning of the huge range of forms and techniques that credit card frauds might take. Some want for nothing in exchange for goods, while others are intended to profit from accounts. With the rise and widespread use of the Internet, the market for credit card payments has increased tremendously over time. The majority of companies and sectors have moved their operations online in order to offer e-commerce, connection, and ease of access to improve productivity and accessibility for their clients. Therefore, it is the obligation of card issuers or the banks to come up with a practical solution to lower any costs that may arise from the problems. Data mining is one approach to solving these problems.

The following figure 2 shows the fraud identification scheme proposed:

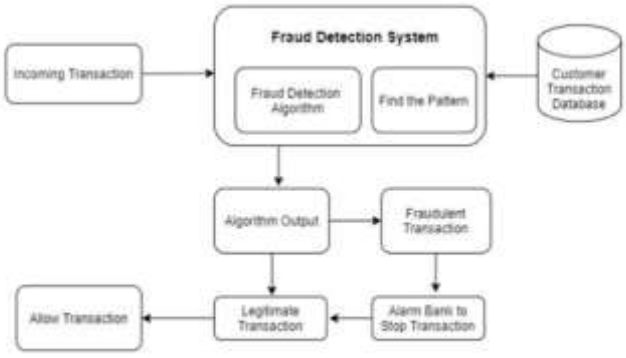

Figure 2: Proposed Fraud detection System

2. Literature Survey

In paper [1] They put into practise a number of the ML techniques suggested in this paper, including the DT, RF, ET, XGB, LR, and SVM. Each proposed algorithm was also combined with the AdaBoost method to improve classification accuracy and address the problem of class imbalance that is prevalent in the European credit card fraud dataset. Additionally, a comparison analysis between the techniques in this work and current framework for detecting credit card fraud was done. For instance, the accuracy of the DT-AdaBoost, RF-AdaBoost, ET-AdaBoost, and XGB-AdaBoost, respectively, was 99.67%, 99.95%, 99.98%, and 99.98%. The XGB-AdaBoost and ET-AdaBoost both received MCCs of 0.99 for the quality of their categorization results. These results proved that applying the AdaBoost algorithm improves the suggested ML techniques. A highly skewed synthetic credit card fraud dataset was used to validate the methodology suggested in this study, and the outcomes were excellent. The ETAdaBoost, for instance, achieved an accuracy of 99.99% and an MCC of 0.99.

In paper [2] Machine learning has become increasingly important in recent years for spotting credit card fraud, but class inequality has posed considerable difficulties. An effective method for detecting credit card fraud was suggested in this research. First, a balanced dataset was produced using the SMOTE-ENN method. Second, utilising the LSTM neural network as the base learner in the AdaBoost method, a strong deep-learning ensemble was created. The suggested LSTM ensemble with SMOTE-ENN data resampling outperformed previous benchmark algorithms and cutting-edge techniques by achieving a sensitivity of 0.996, a specificity of 0.998, and an AUC of 0.990 using the well-known credit card fraud detection dataset. Therefore, combining the boosted LSTM classifier with the SMOTE-ENN data resampling strategy is an effective way to find fraud in credit card transactions.

In paper [3] For the issue that low-frequency users cannot adequately characterise transaction behaviour, a new method for low-frequency user transaction detection is proposed. When compared to existing methods, the suggested method takes into account more than just the transaction behaviour of low-frequency consumers. This leads to the suggestion of a way to evaluate the user's current transaction based on user behaviour and Naive Bayes detection. Additionally, the experiment takes into account the challenge of dividing high-frequency users from low-frequency users. The suggested method's recall is not the highest on all data sets, but experiments show that it has always had lower disturbance rates than other models while still having higher precision and F1. It demonstrates that in the identification of fraud transactions and the detection of low-frequency user transactions

In paper [4] We discovered that by having the retailer make transactions that are highly likely to be fraudulent, credit card fraud can be decreased. The merchant assistance function is addressed by the system's performance rate, which is multiplied by using the data mining technique and random forest algorithm. Although there are now similar schemes on the market, this programme is created to focus specifically on the retailer side of the industry and would benefit the retailer by lowering the fees the merchant must bear in the event that a transaction turns out to be fraudulent. When handling a transaction, the system relies on the merchant's sales and sale data as well as data from the payment gateway, which makes it harder to spot fraud. The plan could improve its ability to thwart fraud.

In paper [5] False statements, claims, documents, or medical problems are used in healthcare fraud to obtain an unauthorised benefit. Attempts to solve the following issues: Our main worry is the strain on the healthcare system. When doctors or other practitioners recommend needless therapies, the systems could suffocate and become overloaded. Such unnecessary procedures could waste resources and result in a lack of medical supplies. Genuine patients are consequently denied access to the necessary healthcare resources or services. Overdiagnosis of diseases can also hurt individuals physically, and in some situations, this can have fatal consequences. For instance, a patient may purposefully receive a cancer diagnosis to attract attention and make it easier to access medical care. However, it might harm the patients physically. The higher premium costs have put a hardship on employers and businesses. Existing research has made tremendous progress in identifying particular fraud categories, but it hasn't succeeded in giving us a standardised method for identifying all forms of healthcare fraud.

3. Data Collection

Due to the general population lacks technology literacy, it's critical to avoid miscommunication between the public and financial institutions like banks in order to reduce the prevalence of fraud. Considering how frequently fraud occurs, this appears to be a major issue. There may be a number of various tactics employed to reduce fraud. In this, our primary goal is developing the best algorithm to prevent fraud and lower the fraud rate. In contrast to conventional methods of gathering credit card data, which include using expiry date and cvv. Make use of fraud detection algorithm to address the shortcomings of fraudulent activities. This paper describes a technique for preventing from frauds using fraud detection algorithms.

4. Methodology

As is well known, the payment gateway is a secure platform for accepting transactions, on which information is provided to confirm the transaction's validity and enable the transfer. However, it is essential to understand how the programme functions in order to spot a scam if a transaction is generated. The proposed programme is envisioned as a component of the payment gateway that would check a transaction for fraud. The fraud detection module will operate in the manner described below: Credit card details, including card number, expiration date, and other information, must be provided by the payment gateway. The merchant will include details like the postal address, the sales number, the delivery date and time, etc. The Fraud Detection Program must receive the proper parameters from the payment gateway.

In order to generate results using the optimal classification algorithm, our proposed application would train itself utilising efficient data mining techniques. Depending on the information, the payment gateway would receive the final result of a transaction, whether it was fraudulent or not. The payment gateway administrator will deliver the verdict and other pertinent information to the merchant in the final user interface report. The random forest method is employed within the framework to analyse the cardholder data combined with the credit card details and the location of the transaction in order to mimic actual transactions. Figuratively 3. Random forest's performance is likely to provide the most accurate indication of the level of fraud.

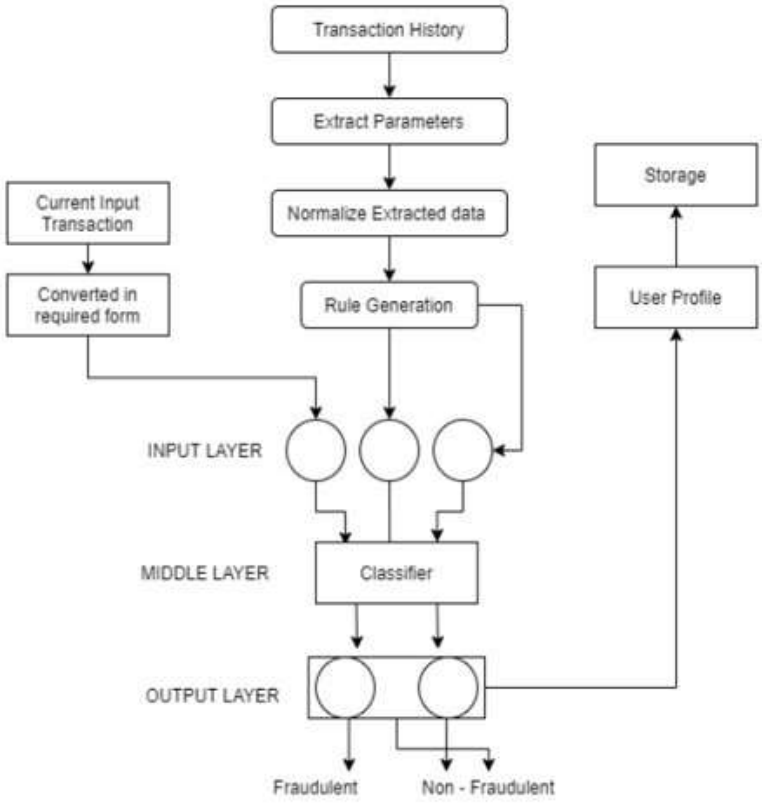

Figure 3: Methodology of the system

5. Results and Discussion

According to experimental findings, the random forest algorithm performs better at accurately predicting the scope of fraud. Figures 4 demonstrate how employing random forest over other algorithms has improved evaluation results for identifying fraudulent transactions.

The proposed system incorporates the algorithm with higher accuracy in identifying fraudulent transactions to enhance performance.

Algorithm/Evaluation Measures	Accuracy	Precision	Recall	F-1 Score
Decision Tree	**83%**	**83%**	**80%**	**81%**
SVM	**86%**	**85%**	**85%**	**85%**
Naive Bayes	**80%**	**85%**	**62%**	**75%**
Random Forest	**88%**	**87%**	**80%**	**88%**

6. Conclusion

The retailer should make purchases that are likely to be fraudulent in order to reduce credit card fraud using the algorithm we built. The system's performance rate more than doubles when the data mining technique and random forest algorithm are combined, addressing the merchant assistance

function. While this method is being developed, it would specifically target the retailer side of the industry and work to their advantage by reducing the expenses the retailer would have to bear in the event that a transaction turned out to be fraudulent. The system relies on the merchant's sales and sale data as well as data from the payment gateway to process a transaction, which makes it harder to spot fraud. The plan can be improved to be more successful in preventing fraudulent transactions.

7.REFERENCES

[1]. S.P. Maniraj, A. Saini, S. Ahmed, and S. Sarkar, ‘‘Credit card fraud detection using machine learning and data science,’’ Int. J. Res. Appl. Sci. Eng. Technol., vol. 8, no. 9, pp. 3788–3792, Jul. 2021.

[2]. N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, ‘‘An efficient credit card fraud detection model based on machine learning methods,’’ Int. J. Adv. Sci. Technol., vol. 29, no. 5, pp. 3414–3424, 2020.

[3]. B Meenakshi, J. B.. S Gayathri, "Credit Card Fraud Detection Using Random Forest", vol. 06, no. 03, March 2019, ISSN 2395-0072

[4]. S. Khatri, A. Arora, and A. P. Agrawal, ‘‘Supervised machine learning algorithms for credit card fraud detection: A comparison,’’ in Proc. 10th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence), Jan. 2020, pp. 680–683.

[5]. A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, ‘‘Real-time credit card fraud detection using machine learning,’’ in Proc. 9th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence), Jan. 2019, pp. 488–493