WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

M.L BASED CREDIT CARD SYSTEM FOR FRAUD DETECTION

Aman Kumar Singh

Raj Kumar Goel Institute of Technology, India

ABSTRACT :

Over the past few years, instances of credit card fraud have been recognized as a growing problem, leading to significant financial losses for individual credit card users, merchants, and banks. The identification of fraud is widely regarded as highly successful through the utilization of machine learning. This manuscript critically examines various fraud detection methods employing machine learning, conducting a compar ison based on performance metrics including accuracy, precision, and specificity.. The proposed Fraud Detect System (FDS) employs a supervised Random Forest algorithm, resulting in increased accuracy in the detection of credit card fraud. Additionally, a learning approach is utilized rank the alerts, effective addressing the issue of concept drift in fraud detection within the proposed system.

Index Terms: Concept drift is observed in the context of credit card fraud using M.L, particularly in the utilization of the Random Forest algorithm.

1. INTRODUCTION

Credit card fraud, involving payment cards like credit cards as an illegal source of funds in transactions, is considered a major problem. Goods and funds are illicitly obtained through fraud, which serves as an illegal means to acquire products or unauthorized funds. Identifying such fraudulent activities poses challenges and may jeopardize businesses and organizations. In real world, all financial transactions cannot be checked by investigators within the Fraud Detection System (FDS) [1]. Here, all approved transactions are monitored by the Fraud Detection System, and the most suspicious ones are alerted. These alerts are verified by investigators, who provide feedback to the FDS regarding the authorization status of the transactions. The process of verifying all alerts every day is a time-consuming and costly endeavor. Consequently, investigators can only verify a limited number of alerts each day, leaving the remaining transactions unattended until customers recognize and report them. fraudulent. Additionally, the techniques used for fraud and changes in cardholder spending behavior occur over time. This alteration in credit card transactions termed as concept of drift, making it challenging to identify credit card fraud most of the time. Machine Learning is regarded as one of the most successful techniques for fraud detect, It employs a classification and regression approach to identify fraudulent activity in credit cards. The machine learning algorithms are divided into two categories: supervised [14][18]and unsupervised [16] learning algorithms. Supervised learning algorithms use labeled transactions for training the classifier, while unsupervised learning algorithms employ peer group analysis [23], grouping customers based on their profiles and identifying fraud based on customers' spending behavior.

RELATED WORK

A variety of supervised and unsupervised learning algorithms are utilized in the detection of fraudulent activities in credit card transactions.. Below are descriptions of some important ones. In a paper proposed by the author [1], proper performance measures used for fraud identification are first explained. A novel learning technique is crafted by the authors to tackle the challenges of concept drift and verification latency, and challenges related to class imbalance.The effect of these issues on true credit card transactions is also demonstrated in the paper.

In the second paper [2], the authors present two varieties of classifiers employing random forests. presented by the authors to train the behavioral features of transactions. The two random forests are compared, and their performance on fraud identification in credit cards is analyzed.

In paper [3], a Fraud Detection System (FDS) for credit cards is presented using Artificial Neural Network and Logistic Regression. The system monitors each transaction separately using a classifier, which generates a score for each transaction and labels it as a legal or illegal transaction.
The paper introduces a decision tree approach / proposes a decision tree method. aiming to decrease overall misclassification costs and select splitting properties at each node. The decision tree method for fraud identification is compared with other models, demonstrating its performance using measures like accuracy and genuine positive rate.

In paper [5], an FDS for credit card transactions is developed using support vector machines and decision trees. Seven alternative models are built using support vector machines and decision trees, and their performance is compared using accuracy as a measure. The study shows that as the size of the training dataset increases, the number of frauds detected the number of frauds detected by support vector machines is lower than those recognized by the decision tree method.
In the sixth citation, the author presents a fraud detection system employing a Naive Bayes K-Nearest Neighbors methodology. The key goal of the suggested system is to improve accuracy. The Naive Bayes Classifier anticipates probabilities of fraud in transactions, and concurrently, the KNN classifier assesses the proximity of undefined sample data to the kth training dataset. Both components contribute to the overarching objective components contribute to the overall effectiveness of the system. classifiers are compared, demonstrating their different performances for the given dataset.

The issue of concept drift is faced by most predictive models used for detecting fraud in credit card transactions. In paper [7], The paper introduces two Fraud Detection Systems (FDS) based on sliding window and ensemble learning, highlighting the necessity to train classifiers independently using feedback and delayed samples. The results from both systems are combined to improve the precision of alerts in the FDS.. It is demonstrated. The author emphasizes the separation of handling feedback and delayed samples as a strategy to tackle the challenge of concept drift.

COMPARISON

Table 1 presents a comparative analysis of the effectiveness of all learning algorithms employed for detecting fraud in credit card transactions. The assessment is grounded in metrics such as accuracy, precision, and specificity.

TABLE 1

COMPARISON OF MACHINE LEARNING TECHNIQUES

Classifiers	Metrics		
	Accuracy	Precision	Specificity
Random Forest	0.9	0.98	0.985
Logis Regres	0.9	0.96	0.96
Knn	0.9	0.41	0.92
svm	0.93	0.732	0.919
Decisio	0.90	0.910	0.946
Naive Bayes	0.93	0.504	0.963

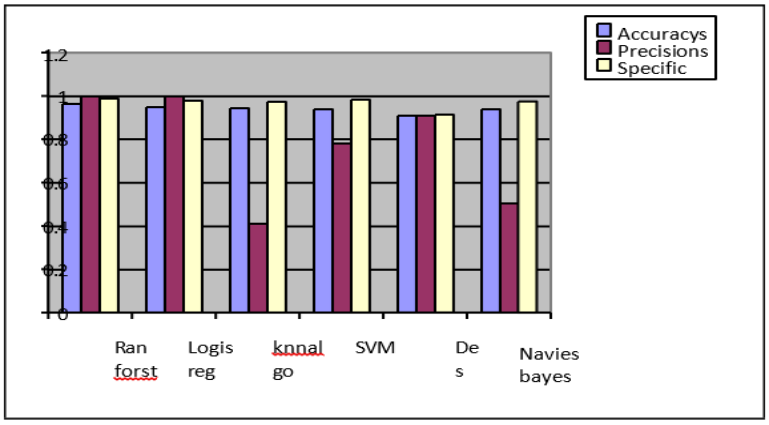

Fig1 The evaluation is being conducted on the performance metrics of accuracy, precision, and specificity for all classifiers..

Table 1 reveals a notable superiority in the accuracy of Random Forest compared to other learning algorithms. Figure 1 visually demonstrates that Random Forest surpasses others in precision, accuracy, and specificity, encompassing Logistic Regression, SVM, Decision Tree, Naive Bayes, and KNN following in descending order. Consequently, the utilization of the proposed system with Random Forest is anticipated to yield superior accuracy, particularly with a larger volume of training data.

Proposed system

Credit cards serve diverse purposes in contemporary society.
Correspondingly, the incidence of fraud in credit card transactions has experienced an upward trajectory in recent years. Annual financial losses of considerable magnitude result from illicit credit card transactions. Fraud can take
varied forms and may have limitations. Consequently, there is a pressing necessity to address the challenges associated associated with fraud detection in credit cards. Additionally, the emergence of new technologies opens up innovative opportunities for criminals to engage in fraudulent activities. In addressing this challenge, a proposed system aims to identify and prevent fraud in credit card transactions is envisaged, employing Machine Learning (ML) techniques. This system aims to furnish investigators with concise and dependable fraud alerts, mitigating the impact of fraudulent activities.

4;(1)objectives

The following main objectives will be achieved by the proposed system:

- The training of the model will involve utilizing feedback and delayed samples, combining their probabilities to recognize alerts.
- A machine learning approach will be deployed to tackle the challenges of concept drift and class imbalance.
- The development of a learning-to-rank strategy is planned to enhance the precision of alerts.
- Introducing performance metrics that are relevant in real- world Fraud Detection Systems (FDS).

Putting forth a Fraud Detection System (FDS), the central focus revolves around a data-driven model and the implementation of a learning-to-rank technique. Additionally, the system underscores the interaction with alert feedback, scrutinizing how recent supervised samples are delivered. The visual representation of the proposed system is depicted in Figure 2.

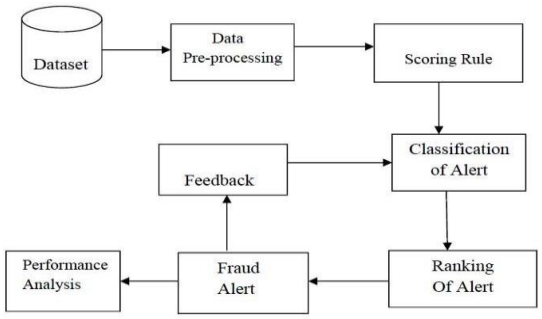

Fig. 2 The Block Diagram of the Proposed System is shown.

Modules

The system suggests the inclusion of the subsequent modules, covering functional necessities:a. Data Preprocess, Score rule, alert distribution, alerting rank, performance analysis component.

Data Preprocess

The chosen data might not align with the required format, potentially existing in a file format when a relational database format is preferred, or vice versa.

Cleaning involves the removal or rectification of missing data. The dataset may contain incomplete records or those with null values, necessitating their elimination**.**

c. Sampling: Addressing the issue of unbalanced class distribution in credit card transactions, where the number of frauds is less than overall transactions, a sampling method is employed as a resolution.

Scoring Rule

The score assigned to the transaction corresponds to the percentage of fraud. This module allocates a score by comparing the recent transaction pattern with the historical transaction pattern of the cardholder. If the score exceeds a certain threshold, the transaction is deemed fraudulent. If the transaction is deemed suspicious, subsequent actions are halted; otherwise, it advances to the subsequent module.

Classification of Alert

In this situation, a machine learning model is used to train and update data, incorporating feedback and delayed samples.

The classifier undergoes separate training sessions with feedback and delayed samples, and their probabilities are combined to detect alerts. Transactions with high probability levels are considered, and alerts are subsequently generated, leading to a limited number of alerted transactions reported to investigators.

Rank alert

Within this module, every alert undergoes a ranking process contingent on the accuracy of the security question. Security questions are generated whenever a transaction is flagged as suspicious. The ranking of alerts is determined by their likelihood. If an alert exhibits a higher probability than others, it is included in a queue, and the fraudster's location is monitored. This functionality enhances the user-friendliness of the system and facilitates the process of filing a complaint against fraud.

5. Conclusion

This paper conducts an extensive examination of various machine learning algorithms tailored for the detection of fraud in credit card transactions. The efficacy of these techniques is assessed using metrics such as accuracy, precision, and specificity. The chosen supervised learning approach, A Random Forest classifier is utilized to classify alerts as either fraudulent or authorized. This classifier undergoes training incorporating feedback and delayed supervised samples, followed by the aggregation of probabilities for alert detection. Furthermore, the proposal introduces a learning-to-rank strategy to prioritize alerts, addressing challenges associated with class imbalance and concept drift. Future efforts will include applying semi-supervised learning methods for alert classification within the Fraud Detection System (FDS).