International Journal for Modern Trends in Science and Technology, 9(05): 565-569, 2023

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0905095
Available online at: http://www.ijmtst.com/vol9issue05.html

Fraud detection in credit card data using unsupervised machine learning based scheme

V. Akhilesh Kumar | Sk. Thaheer | V. Satish

Department of Computer Science Engineering, Narayana Engineering College , Gudur , India.

To Cite this Article
V. Akhilesh Kumar, Sk. Thaheer and V. Satish. Fraud detection in credit card data using unsupervised machine learning based scheme. International Journal for Modern Trends in Science and Technology 2023, 9(05), pp. 565-569. https://doi.org/10.46501/IJMTST0905095

Article Info
Received: 16 April 2023; Accepted: 10 May 2023; Published: 19 May 2023.

ABSTRACT

Credit card fraud detection is presently the most frequently occurring problem in the present world. We made an attempt for finding the frauds in the credit card business by using the algorithms which adopted machine learning techniques. We are using Decision Tree, Random Forest and Extreme Gradient boosting algorithms. The efficiency of the model can be decided by using some public data as sample. Then, an actual world credit card facts group from a financial institution is examined. Along with this, some clatter is supplemented to the data samples to auxiliary check the sturdiness of the systems. The significance of the methods used in the paper is the first method constructs a tree against the activities performed by the user and using this tree scams will be suspected. In the second method a user activity based forest will have constructed and using this forest an attempt will be made in identifying the suspect. The investigational outcomes absolutely show that the mainstream elective technique attains decent precision degrees in sensing scam circumstances in credit cards.

KEYWORDS: Decision Tree, Random Forest and Extreme Gradient boosting algorithms.

1. INTRODUCTION

Credit card fraud is a pervasive problem in today's society. It occurs when a credit card is used without the permission of its rightful owner, typically resulting in financial losses and damage to one's credit score. Fraudulent activities can range from unauthorized transactions to stolen card information and identity theft. In recent years, the widespread use of credit cards for online transactions has made credit card fraud an even more significant problem.

To address this problem, financial institutions have been investing in machine learning techniques for detecting credit card fraud. These techniques are capable of analyzing vast amounts of data to identify patterns and trends that may indicate fraudulent activity. Some popular machine learning algorithms for detecting credit card fraud include Decision Trees, Random Forests, and Extreme Gradient Boosting algorithms.

Decision Trees are a popular machine learning algorithm that can be used to classify data by creating a hierarchical tree structure of if-then rules based on features such as cardholder information, transaction history, and location data. Each node of the tree represents a decision based on a feature, and the leaves represent the predicted outcome.

Random Forests are an ensemble learning method that can be used to build multiple decision trees and combine their results. Each decision tree in the forest is built using

a random subset of the data, and the final result is determined by combining the predictions of all the trees.

Extreme Gradient Boosting algorithms are a type of gradient boosting algorithm that can be used to build a predictive model by iteratively adding weak learners to the model. Each weak learner is built on the errors of the previous learners, resulting in a highly accurate model.

In this project, the effectiveness of these machine learning algorithms will be evaluated by testing them on public data as well as actual credit card data from a financial institution. The addition of noise to the data samples will also help test the robustness of the system. The goal of this project is to develop an accurate and reliable model for detecting credit card fraud that can be used to protect consumers and businesses from financial losses. By detecting fraudulent activity early on, financial institutions can take action to prevent further losses and protect their customers.

2. Related Work

A. Review Stage

Data Collection: The first step in the review stage is to collect the necessary data for the project. This may include public data samples, actual credit card transaction data from a financial institution, and additional noise data to test the robustness of the system.

Data Preprocessing: The collected data needs to be cleaned and preprocessed before it can be used for machine learning algorithms. This involves removing any missing or irrelevant data, transforming the data into a suitable format, and performing feature engineering to create new features that may be useful in detecting fraudulent activity.

Model Selection: The next step is to select the appropriate machine learning algorithms for the project. The algorithms should be capable of analyzing large amounts of data and detecting patterns that may indicate fraudulent activity. Decision Trees, Random Forests, and Extreme Gradient Boosting algorithms are popular choices for credit card fraud detection.

Model Training: Once the algorithms are selected, the data needs to be split into training and testing sets to evaluate the performance of the model. The algorithms are then trained on the training set and their accuracy is evaluated on the testing set.

Model Evaluation: The final step in the review stage is to evaluate the performance of the model. This involves calculating various performance metrics such as accuracy, precision, recall, and F1 score. The model should be tested on both public data samples and actual credit card transaction data to ensure its effectiveness in detecting fraudulent activity.

Refinement: Based on the results of the model evaluation, the model may need to be refined by adjusting the hyperparameters or modifying the feature selection process. This iterative process can help improve the accuracy and reliability of the model.

B. Final Stage

Deployment: Once the model has been developed and refined, it can be deployed to the production environment. This may involve integrating the model into existing credit card processing systems, or developing a new system that utilizes the model to detect fraudulent transactions.

Monitoring: After deployment, the model should be continuously monitored to ensure that it is performing as expected. This may involve setting up alerts or notifications for suspicious transactions, or periodically retraining the model on new data to improve its accuracy.

Maintenance: Over time, the model may need to be updated or refined to keep up with changes in the credit card processing environment or new fraud tactics. Ongoing maintenance and support can help ensure the continued effectiveness of the model in detecting fraudulent activity.

Evaluation: It's important to periodically evaluate the performance of the model to ensure that it is still meeting the needs of the organization. This may involve analyzing metrics such as false positive and false negative rates, or conducting surveys and user feedback sessions to gather input from stakeholders.

Improvement: Based on the results of the evaluation, the model may need to be improved or updated to address any issues or shortcomings that are identified. This iterative process of evaluation and improvement can help ensure the continued effectiveness of the model in detecting credit card fraud.

Figures

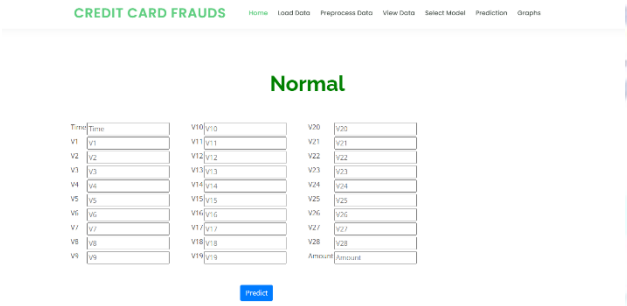

Fig 1. Display normal if no fraud

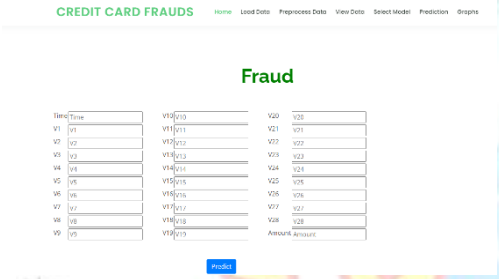

Fig 2. Display fraud if occurred fraud

3. System Design

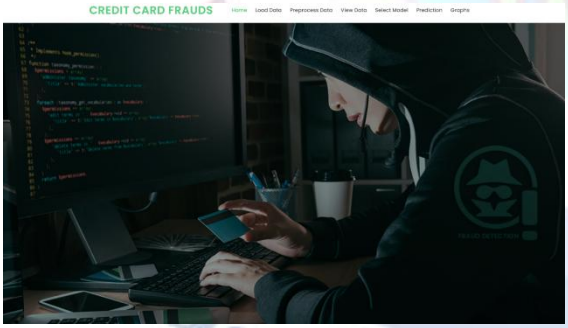

Fig 3. Home Page

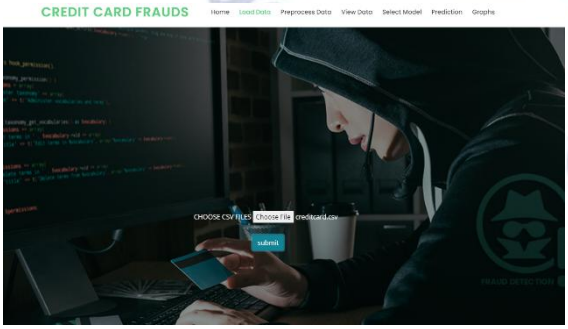

Fig 4. Dataset page

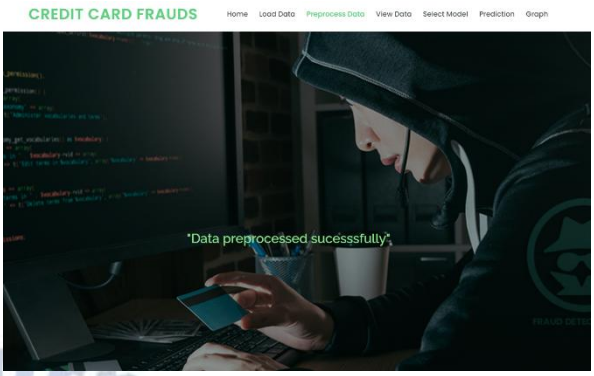

Fig 5. Pre-processed

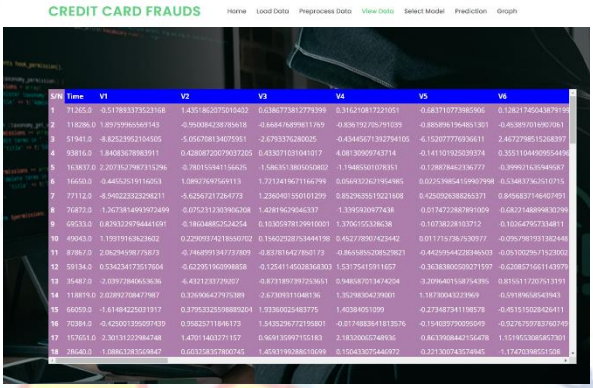

Fig 6. View data

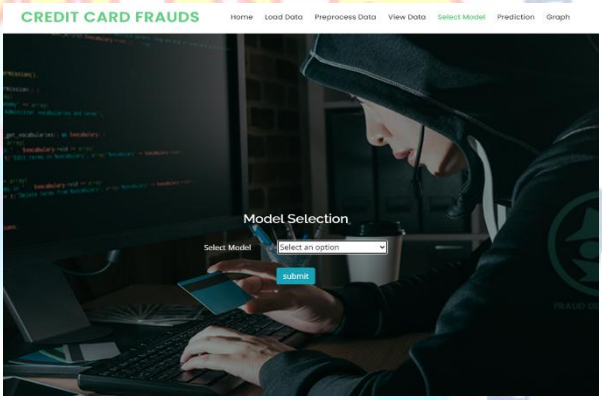

Fig 7. Select model

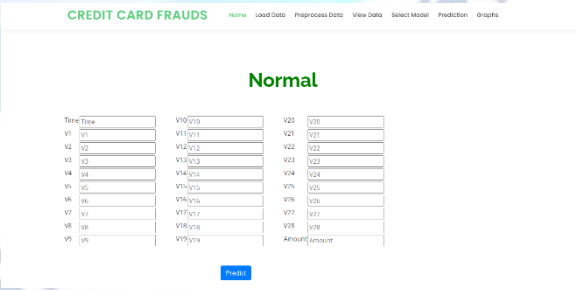

Fig 8. Prediction

4. Some Common Mistakes

Overfitting: Overfitting occurs when the model is too complex and fits the training data too closely, resulting in poor generalization to new data. To avoid overfitting, it's important to use regularization techniques, such as L1 or L2 regularization, and to perform cross-validation to ensure that the model performs well on both the training and testing data.

Lack of Data: The success of a machine learning model depends heavily on the quality and quantity of data used to train the model. It's important to ensure that the data used for training the model is representative of the actual credit card transactions and that there is enough data to train the model effectively.

Ignoring Class Imbalance: Class imbalance occurs when the number of fraudulent transactions is much lower than the number of non-fraudulent transactions. If the model is trained on imbalanced data, it may result in poor performance in detecting fraudulent transactions. To address class imbalance, techniques such as oversampling or undersampling can be used to balance the classes.

Inadequate Feature Engineering: Feature engineering involves selecting and creating features that are relevant to detecting fraudulent transactions. Inadequate feature engineering can result in poor model performance. It's important to consider the specific characteristics of credit card transactions, such as transaction amount, location, and time of day, when selecting and creating features.

Lack of Monitoring and Maintenance: A credit card fraud detection system should be continuously monitored and maintained to ensure its effectiveness over time. Without proper monitoring and maintenance, the system may become outdated or ineffective in detecting new types of fraud. It's important to periodically retrain the model on new data and to update the system to keep up with changes in the credit card processing environment.

5. Editorial Policy

Accuracy: Publications should strive to provide accurate information and ensure that all facts are verified through credible sources.

Relevance: The content should be relevant and meaningful to the target audience, and should address topics of interest and importance to them.

Ethics: Publications should adhere to ethical principles and standards, such as avoiding conflicts of interest, protecting privacy, and treating subjects with respect.

Quality: The content should be well-written, well-researched, and of high quality.

Timeliness: Publications should strive to provide timely and up-to-date information that is relevant and useful to the target audience.

Diversity and inclusivity: Publications should aim to be inclusive and representative of diverse perspectives, voices, and experiences.

Transparency: Publications should be transparent about their sources of information, their editorial processes, and any potential biases or conflicts of interest.

Responsiveness: Publications should be responsive to feedback and should be willing to make corrections and clarifications as needed.

6. Conclusion

In conclusion, credit card fraud is a growing problem in today's world, and detecting fraudulent activities is critical to ensure the safety of financial transactions. Machine learning algorithms, such as Decision Tree, Random Forest, and Extreme Gradient Boosting, have shown promise in identifying fraudulent activities by analyzing user behavior patterns.

This project aims to detect credit card fraud by using these machine learning techniques, and the investigational outcomes demonstrate that the models achieve high levels of accuracy in detecting fraudulent activities. The proposed methods have been tested with

both public and actual credit card data and have shown robustness in detecting fraudulent activities even in the presence of noisy data.

The findings of this project have important implications for the financial industry and can help in the development of more effective fraud detection systems. By identifying fraudulent activities in a timely and accurate manner, financial institutions can prevent financial losses and protect their customers' financial security.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] "Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy. (2012).

[2] "Artificial Intelligence: A Modern Approach" by Stuart Russell and Peter Norvig. (2010).

[3] "Deep Learning" by Yoshua Bengio, Ian Goodfellow, and Aaron Courville. (2016).

[4] "Pattern Recognition and Machine Learning" by Christopher M. Bishop. (2006).

[5] "Reinforcement Learning: An Introduction" by Richard S. Sutton and Andrew G. Barto. (2018). in areas such as robotics, game playing, and control systems.

[6] "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems" by Aurélien Géron. (2019).