WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Credit Card Fraud Detection Project Using Machine Learning

Monisha S

Assistant Professor, *Sacred Heart Arts and Science College, Perani, Villupuram 605 651, Tamilnadu*

A B S T R A C T

Credit card fraud is an easy and friendly target. E-commerce and many other online sites have expanded online payment methods, increasing the risk of online fraud. Increasing fraud rates, researchers have started using various machine learning methods to detect and analyze fraud in online transactions. The main objective of the paper is to design and develop a new fraud detection method for Streaming Transaction Data to analyze the details of customers' past transactions and extract behavioral patterns. Where cardholders are grouped into different groups based on their transaction amount. Then, using a sliding window strategy [1] to aggregate the transactions made by cardholders from different groups so that patterns of individual group behavior can be extracted. Later, different classifiers [3],[5],[6],[8] are trained on the groups separately. And then the classifier with better ranking can be selected as one of the best fraud prediction methods. So the following is a feedback mechanism to solve the concept drift problem

Keywords: Credit, Debit, Online, Card, Fraud, Resulting, Percentage, Fraud and Detection

1. Introduction

When the world was under lockdown and movement was restricted to an absolute state of emergency, millions of people were introduced to the world of online shopping. The convenience of online shopping has helped e-commerce platforms record historic sales. While this has happened, it is no surprise that the rate of online financial fraud has also increased dramatically. Online credit and debit card fraud cases saw a historic increase of 225 percent during the COVID-19 pandemic in 2020 compared to 2019. According to the NCRB report, the number of credit and debit card fraud cases in 2020 was 1,194 compared to 367 in 2019 .According to data revealed by the Reserve Bank of India (RBI) in response to an RTI request, an average of 229 bank frauds were committed every day in India resulting in transactions worth Rs. 1.38 million crowns are taken. The rate of return on the same was also not impressive.

1.1 What is Credit Card Fraud Detection?

Credit card fraud is the term used to refer to unauthorized access to a payment card, such as a credit or debit card, to pay for the use of services or goods. Hackers or fraudsters can get confidential card information from unsecured websites. When fraudsters compromise an individual's credit/debit card, everyone involved in the process is affected, with the individual's confidential data leaked to both the company that issued the credit card (usually the bank) and to the merchant who finalizes the purchase transaction. It is therefore extremely important to identify fraudulent transactions in the first place. Companies such as financial institutions and e-commerce are taking strong action to flag fraudsters who enter the system.

There are various advanced techniques to create assembly models that are more powerful and robust.

1.2 Stacking

In this stacking technique, base models are trained on the data and predictions from these models are used to train the final classifier and can make final predictions on the test data.

1.3 Boosting

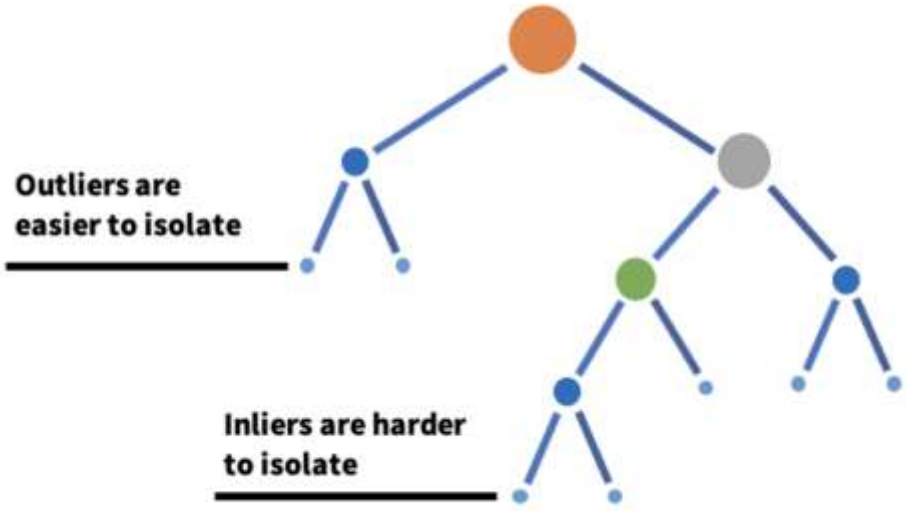
In this ensemble technique, a random sample of data is selected from the training data to train the first weak learner such as a decision tree model. Then, the data points that the first model incorrectly predicts are combined with another random sample from the training data, and a second weak learner is trained using this data. Similarly, a third weak learner is trained on data where the second model predicts incorrectly. Each data point is initially assigned equal weight, but as weak learners are trained, the weight of data points that are incorrectly predicted increases, so the next model puts more emphasis on them. Models like AdaBoost, XGBoost, LightGBM, etc. use this technique in training.

1.4 Bagging

In this ensemble technique, several weak learners are trained on random samples of data from the training data and then their output is combined into a final prediction. Each model is trained on a different data sample and is independent of each other, allowing all models to learn different data features. The RandomForest algorithm is based on this technique.

Unsupervised algorithms for credit card fraud detection

These algorithms identify patterns in data that are neither labeled nor classified. Therefore, no external training instructions are provided to the model, allowing unsupervised algorithms to perform more complex processing tasks than supervised systems. But the lack of labels also makes unsupervised learning methods a bit unreliable, as they might identify unforeseen categories for unusual data. Algorithms understand the underlying data and extract information from it to create groups of data.



2. How does credit card fraud work

A credit card is one of the most used financial products for online purchases and payments such as gas, groceries, TV, travel, shopping bills and so on, because the funds are not available in that case. The most valuable are credit cards that provide various benefits in the form of points and at the same time use them for various transactions. Several categories of credit card fraud are prevalent today:

Lost/Stolen Cards: People steal credit cards from the post office and use them illegally in the owner's name. The process of blocking stolen credit cards and reissuing them is a problem for both customers and credit card companies. Some financial institutions keep credit cards blocked until it is verified that the rightful owner has received the card.

Card abuse: A customer purchases goods and items on a credit card but has no intention of repaying the amount charged by the bank. These customers stop answering calls when the due date is approaching. Sometimes they even declare bankruptcy - this type of fraud leads to millions in losses every year.

Identity Theft: Customers use illegitimate information and may even steal real customer information to apply for a credit card and then misuse it. In such cases, even blocking the card cannot prevent the credit card from falling into the wrong hands.

Merchant Abuse: Some merchants report illegal transactions (which never happened) in order to launder money. In order to carry out these illegal transactions, the legal information of real credit card users is stolen in order to create replica cards and use them for illegal work.

Many old-school traditional techniques have been used since time immemorial to detect credit card fraud like CVV verification, geolocation tracking, IP address verification, etc. But with time, criminals are using more advanced techniques to commit crimes. not all of them can be prevented using traditional methods alone. In today's world, millions of transactions are processed every second, which exceeds human intelligence to process all the data to identify fraudster behavior patterns. This is where machine learning credit card fraud detection plays a vital role.

Financial institutions increasingly depend on automated machine learning systems to make intelligent decisions and protect businesses from significant losses. These measures play a significant role in reducing the risk of conducting online transactions. Although machines may not be as intelligent as humans and may also need some supervision, the advantage lies in the speed of data processing and calculations. Machines can also identify and remember more patterns in vast amounts of data compared to humans.

3. Algorithm

3.1 Isolation forest

This is an unsupervised algorithm that was built using multiple decision trees, just like Random Forest. The algorithm selects a random sample of data for a given data set and splits the data based on a randomly selected function from the data set. A threshold between the minimum and maximum value of the selected function is randomly chosen to divide the data.4. Online license transfer

3.2 Local outlying factor

The local outlier algorithm identifies such According to the LOF algorithm, those with low density are outliers. Based on the reachability distance calculated for the k-nearest points, the local reachability density of all data points (the LRD is calculated). The LOR, or local outlier factor, is calculated as the ratio of the average LRD of a data point's nearest neighbors to the LRD of a given data point.

3.3 One SVM class

But what if we only have one class and want to find out if a new data point is from that class or an outline. A one-class SVM is trained unsupervised instead of the standard SVM for binary classification, which uses a supervised approach. The hyperparameters of the model are also different from the standard SVM model.

3.4 Automatic encoders

An autoencoder is a special kind of feedforward neural network in which the inputs and outputs of the model are the same. The input data given to the model is compressed to a lower dimension so that only the most important information is extracted from the data. The data is then reconstructed from it. This lower dimensional data representation is also known as latent space representation. The process of converting data from a higher dimension to a lower dimension is known as encoding. The process of reconstructing the data is known as decoding. The loss function compares the input with the reconstructed input, optimized using a cost function during model training. They are lossy, meaning they should not be expected to regenerate the input 100% accurately. Some or other data is lost during the dimension reduction process.

3.5 Controlled credit card fraud detection algorithms

These algorithms use labeled data to train a model. In other words, we can say that for a given set of features, the model learns to predict the target value. The term "supervised" is derived from the idea that the model learns from a training data set under the supervision of a teacher. The model learns while optimizing for the maximum possible correct predictions for the validation data. Obtaining labeled data can sometimes be complex and may even require additional human resources and costs.

3.6 K-nearest neighbor

It is one of the simplest yet most effective supervised machine learning algorithms used in industry. It works on the simple idea that surrounding data is likely to be similar. The k value passed to the model calculates the distance of the data point from all other data points and identifies nearest k neighbors based on the distance. Euclidean distance is primarily used for this task. If a new data point comes in and is closest to the set of data points labeled "class 1", that will be our predicted class.

3.7 Vector Machine support

SVM is another popular supervised algorithm and works well for both regression and classification problems. The basic idea is to identify a hyperplane in the N-dimensional feature space that can be used to segregate data points into their target classes. The best hyperplane for the SVM is chosen by maximizing its distance from the nearest data point on either side of the plane. It is also known as hard edge. These closest data points on either side of the hyperplane are the support vectors.

3.8 File models

Ensemble models is a machine learning modeling approach that combines the power of multiple models, also known as weak learners, to create a robust and accurate model. The main idea behind this approach is to use the best of both worlds. It may happen that a single model may not be sufficient to perfectly learn a classification task, as each algorithm has its own limitations. But if we train multiple models and combine the results of each model to predict the class, you can eventually increase the overall performance of the model. There are many ways to build file models. The most direct approaches include:

Voting Classifier: The class predicted by the maximum number of models will be the final predicted class.

Average: Take the average of the probability classes of all models to calculate the final predicted probability.

Weighted average: If you want to favor a particular model during prediction, assign the highest weight to the model in the weighted average.

3.9 Neural networks

Artificial N consists of three layers: input, hidden, and output. Each layer consists of a small unit where data transformation techniques are used to extract hidden patterns and information from the data. These transformations are known as activation functions. There are different types of activation functions like ReLU, sigmoid, tan-h etc. Neural network uses forward pass and back propagation to train the model. The input is passed through the neural network during a forward pass and classes are predicted for the data. A loss is then calculated to indicate how far our predictions are from the underlying marks. Based on the loss, the model performs backpropagation to minimize the loss and update the weights of each unit in different layers for the next forward pass. Acknowledgements

3.10 Challenges in detecting credit card fraud

The challenges associated with the credit card fraud detection project are primarily the data itself. The data is highly unbalanced, i.e. the number of data marked as fraudulent is much smaller than the number of data marked as non-fraudulent. This makes it extremely difficult to train a model as it tends to overfit for the majority class and underfit for the minority class. Techniques like resampling, undersampling, cost-sensitive learning, etc. can be used to solve this problem. The metrics used for the final model are different from the standard accuracy evaluation metrics, AUC-ROC, etc.

Another prevalent issue is data quality and quantity. Early-stage startups don't have a lot of user history data to train large-scale models, making it difficult to train a robust fraud detection model. Obtaining data from an external third party, such as a credit score, may be a temporary solution to this problem.

4. Conclusion

Fraud is a major problem in the entire credit card industry, which is increasing with the increasing popularity of electronic money transfers. Credit card issuers should consider introducing advanced credit cards to effectively prevent criminal acts that lead to bank account leaks, skimming, counterfeit credit cards, theft of billions of dollars annually, and loss of reputation and customer loyalty. Card fraud prevention and detection methods. Finally, it should be noted that anyone looking for web development services related to financial applications that may be potentially vulnerable to fraud should also immediately seek out a web development provider with significant ML experience.

References

1. Changjun Jiang, et al.“Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism.” IEEE Internet of Things Journal, 5 (2018), pp. 3637-3647 View article CrossRefView in ScopusGoogle Scholar

2. Pumsirirat, A. and Yan, L. (2018). Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine. International Journal of Advanced Computer Science and Applications, 9(1). Google Scholar

3. Mohammed, Emad, and Behrouz Far. “Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study.” IEEE Annals of the History of Computing, IEEE, 1 July 2018, doi.ieeecomputersociety.org/10.1109/IRI.2018.00025. Google Scholar

4. Kuldeep Randhawa, et al. “Credit Card Fraud Detection Using AdaBoost and Majority Voting.” IEEE Access, 6 (2018), pp. 14277-14284 doi:10.1109/access.2018.2806420 View article CrossRefView in ScopusGoogle Scholar

5. Roy, Abhimanyu, et al. “Deep Learning Detecting Fraud in Credit Card Transactions.” 2018 Systems and Information Engineering Design Symposium (SIEDS), 2018, doi:10.1109/sieds.2018.8374722. Google Scholar

6. Xuan, Shiyang, et al. “Random Forest for Credit Card Fraud Detection.” 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018, doi:10.1109/icnsc.2018.8361343. Google Scholar

7. Awoyemi, John O., et al. “Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis.” 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017, doi:10.1109/iccni.2017.8123782. Google Scholar

8. Melo-Acosta, German E., et al. “Fraud Detection in Big Data Using Supervised and Semi-Supervised Learning Techniques.” 2017 IEEE Colombian Conference on Communications and Computing (COLCOM), 2017, doi:10.1109/colcomcon.2017.8088206. Van der Geer, J., Hanraads, J. A. J., & Lupton, R. A. (2000). The art of writing a scientific article. *Journal of Science Communication, 163*, 51–59.