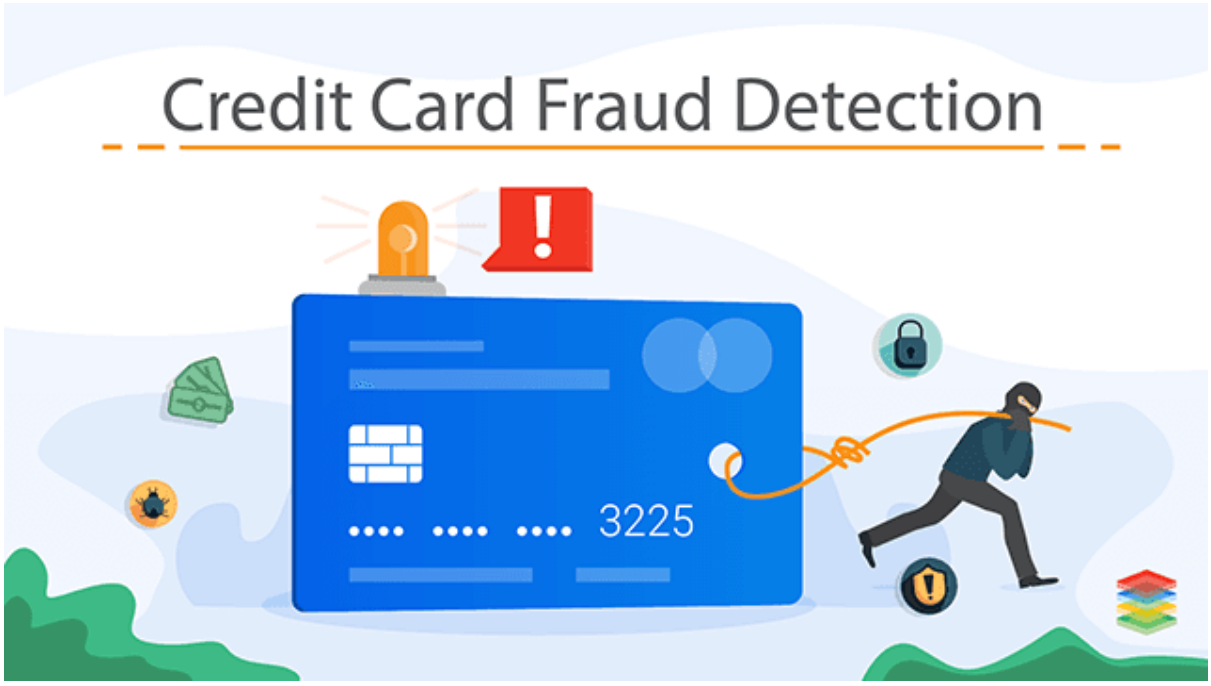

CREDIT CARD FRAUD DETECTION

Submitted By :
Jugal Ramchiary [R21EJ013]

Sudhashu Kumar [R21EJ033]

Guide :

Prof. Pavan Kumar Naik

Introduction

Credit card fraud is a serious and global issue or crime committed by fraud using payment cards such as credit cards or debit cards. The purpose of these fraudsters is to acquire goods without paying or to acquire unauthorized funds from an account. Credit card fraud also gives rise to identity theft. According to some reports and statistics, while the rate of identity theft has been steadily increased by 21 percent in 2008. However, credit card fraud, that crime that leads to ID theft, decreasing as a percentage of all ID theft complaints for many years. Although only 0.1% of cardholders are aware of credit card frauds. These credit card frauds have resulted in huge financial losses as the fraudulent transactions have been large value transactions. In the year 1999, 10 million transactions out of 12 billion turned out to be fraudulent. Also, every 4 out of every 10,000 active accounts are fraudulent. Current fraud detection systems are only able to prevent 1/12th of 1% of all transactions processed which still leads to billions of dollars in losses.

Abstract

In data mining, anomaly detection means to search or scan for a data point, item, or record which do not match or conform to an expected pattern, trend, or other data points in the dataset. So, most of the time these data points or records are considered as defects, outliers, errors, or frauds. Various machine learning anomaly detection algorithms enhance the speed of detecting these outliers. These anomaly detection algorithms are used for detecting invasions while detecting outliers and can also prevent attacks, defects, faults, and so on. Various companies, organizations or institutions adapted and implement these algorithms with a simple yet effective approach for detecting and classifying these anomalies. Machine learning algorithms can learn from data and make predictions based on that data. Since basic machine learning involves learning from data and predict the data but anomaly detection algorithms specifically learn or work on these outliers. It provides an alternative for the detection and classification of anomalies based on an initially large set of features. Anomaly detection or outlier detection is the recognition of unalike data, records, or observations which raise doubts by differing significantly from the majority of the data.

Reference	Method	Method Applied to:	Advantages	Disadvantages
Ghosh and Reilly (1994)	Neural network (restricted coulomb energy algorithm)	Credit card transactions	Increased accuracy and timeliness of fraud detection	Compared to other data mining techniques this method requires a longer training period
Aleskerov, Freisleben, and Rao (1997)	Neural network (gradient descent algorithm)	Credit card transactions	Can handle large commercial size databases	Non-convergence in training
Dorronsoro (1997)	Neural network	Credit card transactions	Real-time fraud detection	Difficulty in determining the optimal size of the hidden layers
Kokkinaki (1997)	Decision tree	Credit card transactions	Simple and easy to implement; reduced misclassifications	Not dynamically adaptive
Ehramikar (2000)	Decision tree	Credit card transactions	Predictive performance was improved by increasing the number of minority instances	Only the decision tree algorithm is experimented upon
Wheeler and Aitken (2000)	Case-based reasoning (Nearest neighbor and probabilistic algorithms)	Credit applications	Model can be easily updated and maintained; robust to missing or irrelevant data	Requires two separate experiments; one to determine the instances to experiment upon, and another to determine the final prediction.
Bolton and Hand (2001)	Outlier detection (unsupervised)	Credit card transactions	Successful in detecting local anomalies and can detect fraudulent behavior in a continuous manner	Treats all accounts equally; does not differentiate between different accounts

Research Gaps

Adversarial Attacks:

The term “adversarial attacks” refers to attacks that attempt to trick machine learning models by manipulating input data. Research could be focused on creating models that are more resilient and less vulnerable to these types of attacks when it comes to credit card fraud.

Real-Time:

Making fraud detection algorithms faster and more accurate, particularly in near real-time situations, is still a work in progress. Research may be focused on creating algorithms that can quickly analyse transactions without sacrificing accuracy.

Behavioral Biometrics:

Exploring the use of behavioral biometrics, such as typing patterns, mouse movements, or interaction behavior, as additional factors for fraud detection.

Cross-channel Fraud Detection:

Integrating information from various channels and sources, such as online transactions, ATM usage, and mobile banking, to create a holistic view of customer behavior and detect fraud across multiple channels.

Conclusion

- Several algorithms have been implemented on the same data set to detect credit card frauds. All the algorithms have been analysed and compared on basis of accuracy on basis of predicting normal cases and outliers or frauds. We implemented a different type of algorithms which include a neural network from deep learning, anomaly detection algorithms like isolation forest, **OneClassSVM**, Local Outlier Factor, supervised algorithms like **DBSCAN**. This was done to attain the best approach for the purpose.

- Upon analyzing we get to know that 3-Layer Neural Network and DBSCAN are spot-on predicting the normal cases with an accuracy of 99% but in the case of predicting the outliers they are not as good as anomaly detection algorithms. Isolation forest and **OneClassSVM** algorithm are giving impressive accuracy of 91% on predicting the outliers but in the case of predicting the normal case they have less accuracy as compared to neural networks. In the case of time taken neural networks and isolation, forest algorithms are very impressive. In the future, this type of algorithm can be used in different cases. For better performance, we can change the properties of the layers of neural networks for better results.