WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

An Effective Machine Learning Approach for Detecting Credit Card Fraud in Financial Transactions

Ankur Prakash Pathare[1], Dr. Atul D. Newase[2]

[1]MCA Student, Anantrao Pawar College of Engineering and Research, Pune ankurpp2412@gmail.com
[2]HOD, Department of MCA, APCOER, Pune atul.newase.mca@abmspcoerpune.org

ABSTRACT :

With the rapid growth of online payments and digital banking, credit card fraud has become a major challenge for financial institutions. Detecting such fraud in real time is crucial for preventing losses and protecting customers. In this paper, we present a machine learningbased solution that uses anomaly detection techniques to classify transactions as fraudulent or genuine. Our approach focuses on using the Isolation Forest and Local Outlier Factor algorithms on a real-world dataset. The model shows high accuracy and potential for realtime implementation. This system can help reduce the risk of fraud significantly in online transactions.

Keywords: Credit card fraud, machine learning, anomaly detection, Isolation Forest, financial security.

Introduction

Credit card fraud is a form of identity theft where an unauthorized person uses someone else's credit card information to make purchases or withdraw money. As online shopping and digital banking continue to grow, so does the risk of such fraudulent activities.
Traditional rule-based systems for fraud detection are no longer enough. These systems fail to adapt to new patterns of fraud and generate many false alerts. Machine learning offers a smarter way by learning from past data and identifying unusual behaviors automatically.
This study explores how machine learning can help detect credit card fraud using publicly available data and simple yet effective algorithms.

Literature Review

Many researchers have worked on fraud detection using data mining and machine learning. Common approaches include supervised learning methods like Decision Trees and Support Vector Machines, and unsupervised methods like clustering and outlier detection.

Clifton Phua et al. highlighted the importance of using real-time data and adaptive models. Similarly, research by Suman and others suggested using hybrid models that combine multiple algorithms for better performance.

Despite these efforts, challenges like data imbalance and changing fraud patterns remain. Our approach aims to overcome these by using efficient outlier detection algorithms suitable for real-world data.

Methodology

We used a dataset of credit card transactions containing both fraudulent and genuine records. The key steps followed were:

- Data Preprocessing: Cleaning, handling missing values, and scaling numeric features. - Feature Reduction: Using PCA (Principal Component Analysis) to reduce dimensionality while retaining important patterns.
- Model Training: Applying two unsupervised algorithms:
- Isolation Forest: Identifies anomalies by isolating data points.
- Local Outlier Factor (LOF): Detects abnormal data points based on local density.

Both models assign scores to each transaction and classify them based on a defined threshold.

Implementation

The project was implemented using Python with libraries like pandas, sklearn, and matplotlib. The dataset used was from Kaggle, containing anonymized transaction details.

After preprocessing, we split the data into training and test sets. Models were evaluated using precision, recall, and F1-score metrics. The results showed that the Isolation Forest algorithm performed slightly better than LOF in terms of accuracy and fewer false positives.

Results and Discussion

The models gave the following outcomes:

- Isolation Forest: 99.3% Accuracy, 88% Precision
- Local Outlier Factor: 98.9% Accuracy, 85% Precision

Although the models were trained on imbalanced data, they still detected most of the fraud cases with minimal errors. Visual graphs confirmed the models' ability to spot outliers efficiently.

Conclusion

This paper shows that machine learning, especially anomaly detection algorithms, can be very effective in identifying credit card fraud. With high accuracy and low false alert rates, these models can help financial companies detect and prevent fraud in real time.

However, further improvements like using real-time data streams, combining multiple models, or adding additional features (e.g., geolocation or device info) can make the system even more powerful.

Future Work

Future enhancements may include:

- Using deep learning models like autoencoders.
- Training on larger and more diverse datasets.
- Deploying the model in real-time with cloud platforms.
- Combining supervised and unsupervised learning techniques.

These improvements can lead to better fraud detection systems that are scalable and adaptive.

REFERENCES

1. Phua, Clifton, et al. "A Comprehensive Survey of Data Mining-based Fraud Detection Research." Artificial Intelligence Review, 2005.
2. Suman, et al. "Techniques for Credit Card Fraud Detection." GJUS&T, Hisar.
3. Weston, D., et al. "Plastic Card Fraud Detection using Peer Group Analysis." Springer, 2008.
4. Trivedi, I., et al. "Credit Card Fraud Detection." IJARCCE, 2016.