Project-1:Anomaly Detection in Financial Transactions Using AI Data Analyst.

College Name:Dr.Ambedkar Institute Technology(Dr.AIT).

Ravi K R **CAN_36049468**

Anomaly Detection in Financial Transactions Using AI Data Analyst

Abstract

Anomaly detection in financial transactions is a critical aspect of preventing fraud and identifying unusual behaviors in real-time. With the rise of machine learning and AI techniques, traditional methods for fraud detection are becoming increasingly sophisticated and efficient. This research explores the application of various machine learning models, such as Isolation Forest, Autoencoders, and One-Class SVM, for detecting anomalies in financial transaction data. The objective of this study is to assess the effectiveness of these models in identifying fraudulent or suspicious transactions while minimizing false positives. Results demonstrate that machine learning models can provide an effective solution for financial institutions to detect and prevent fraud.

1. Introduction

Anomaly detection plays a crucial role in identifying unusual patterns in financial transactions, which could indicate fraud, unauthorized access, or errors. As the financial sector grows and transactions become more complex, there is a growing need for automated systems to identify fraudulent behavior effectively. Traditional methods such as rule-based systems and heuristic models are often limited by predefined patterns and are less effective at detecting new, emerging types of fraud.

This research focuses on leveraging AI techniques, specifically machine learning algorithms, to detect anomalies in financial transaction data. We evaluate the performance of various machine learning models, including **Isolation Forest**, **Autoencoders**, and **One-Class SVM**, to determine their accuracy, precision, and recall in identifying fraudulent transactions.

2. Literature Review

The concept of anomaly detection has been widely researched in the context of various applications, including fraud detection in financial transactions. According to Chandola et al. (2009), anomaly detection involves identifying data points that deviate significantly from the majority of the dataset. This concept has been applied to fraud detection in various sectors, including banking, e-commerce, and insurance. Traditional methods for fraud detection relied heavily on statistical techniques and expert-defined rules, which often failed to adapt to new and evolving fraudulent behaviors.Machine learning models, particularly **supervised** and **unsupervised** techniques, have shown great promise in detecting anomalies. Models like **Isolation Forest** and **One-Class SVM** work well in unsupervised settings where labeled data may be scarce. On the other hand, deep learning techniques like **Autoencoders** have the ability to learn more complex patterns and features from data, providing higher flexibility in identifying fraud.

3. Methodology

3.1 Dataset

The dataset used for this research consists of historical financial transaction data, including details such as transaction amount, time of transaction, user information, and location. The dataset also includes labeled data indicating whether each transaction was fraudulent or not, which is used to evaluate the performance of the models.

3.2 Data Preprocessing

Before applying the machine learning models, the data undergoes preprocessing, which involves several steps:

- **Missing Value Handling**: Missing values are imputed with the median for numerical features and a placeholder category for categorical features.
- **Feature Engineering**: New features are created, such as:
- **Transaction Time**: Extracting time-related features like hour of the day and day of the week.
- **Amount Normalization**: Standardizing the transaction amount using Z-score normalization.
- **User Behavior**: Calculating the frequency of transactions by the same user within a short time window.

3.3 Algorithms Applied

The following machine learning models were applied to detect anomalies in the financial transaction dataset:

- **Isolation Forest**: This model isolates outliers by randomly partitioning the data. Anomalous data points are those that require fewer partitions to separate from the rest of the data.
- **Autoencoder**: A neural network trained to compress and reconstruct data. Transactions with high reconstruction errors are flagged as anomalies.
- **One-Class SVM**: This method identifies a boundary around normal data points. Anything outside this boundary is flagged as an anomaly.

4. Implementation

4.1 Data Preparation

The data is first loaded and cleaned, ensuring that missing or erroneous values are handled properly. Features are selected and transformed to enhance the model's ability to detect patterns.

4.2 Model Training

Each of the selected models is trained on the preprocessed dataset. The models are trained in an unsupervised manner, meaning they learn to identify normal and anomalous transactions without needing labeled fraud data for training.

4.3 Evaluation Metrics

The models are evaluated based on three key performance metrics:

- **Precision**: The ratio of correctly identified fraudulent transactions to all flagged fraudulent transactions.
- **Recall**: The ratio of correctly identified fraudulent transactions to all actual fraudulent transactions.
- **F1-Score**: The harmonic mean of precision and recall, providing a balance between the two.

4.4 Anomaly Scoring and Thresholding

Each model generates an anomaly score for each transaction, and a threshold is set to classify transactions as fraudulent or non-fraudulent. Transactions with scores above the threshold are flagged as anomalous.

5. Results and Discussion

The results show that all three models demonstrated the ability to detect fraudulent transactions. However, each model has its strengths and weaknesses:

- **Isolation Forest** performed well with a low false positive rate but had difficulty detecting subtle anomalies.
- **Autoencoders** achieved the highest precision but had a slightly lower recall, meaning some fraudulent transactions were missed.
- **One-Class SVM** was effective in detecting anomalies, but it required fine-tuning of the kernel parameters to achieve optimal performance.

The models showed that machine learning techniques could significantly improve fraud detection in financial transactions compared to traditional rule-based methods.

6. Conclusion

In conclusion, machine learning models, particularly **Isolation Forest**, **Autoencoders**, and **One-Class SVM**, are effective tools for anomaly detection in financial transactions. These models are capable of identifying fraudulent transactions in large datasets with a reasonable degree of accuracy. While there is still room for improvement, particularly in terms of fine-tuning the models for specific transaction types, the results are promising. Future work could focus on integrating more diverse datasets, real-time processing, and advanced models like **Deep Learning** for further enhancing fraud detection capabilities.

7. Future Scope

There are several directions for future research in anomaly detection for financial transactions:

- **Real-Time Detection**: Implementing real-time fraud detection systems would significantly reduce the time between fraud occurrence and response.
- **Ensemble Models**: Combining multiple models into an ensemble could further improve detection accuracy and reduce false positives.

- **Adaptive Models**: Creating models that can adapt to evolving patterns of fraud over time, using techniques like transfer learning.
- **Explainability**: Enhancing model interpretability so that the reasoning behind flagged transactions can be better understood by financial analysts and stakeholders.

8. References

[1] Amirruddin AD, Muharam FM, Ismail MH, et al (2022) Synthetic Minority Over- sampling TEchnique (SMOTE) and Logistic Model Tree (LMT)-Adaptive Boosting algorithms for classifying imbalanced datasets of nutrient and chlorophyll sufficiency levels of oil palm (Elaeis guineensis) using spectro radio meters and unmanned aerial vehicles. Comput Electron Agric 193:106646–106646

[2] Bolton R, Niveditha DHC, Abarna K, et al (2001) Unsupervised profiling methods for fraud detection. Citeseer, Accessed, URL https://citeseerx.ist.psu.edu/document? repid=rep1&type=pdf&doi=5b640c367ae9cc4bd072006b05a3ed7c2d5f496dG

[3] Dataset ELR (????) URL https://www.kaggle.com/code/netzone/eda-and-fraud- detection/data

[4] Ditzler G, P R, T (2010) URL https://ieeexplore.ieee.org/abstract/document/ 5596764/

[5] Duman E, Buyukkaya A, Elikucuk I (2013) A novel and successful credit card fraud detection system implemented in a Turkish bank. Proceedings - IEEE 13th International Conference on Data Mining Workshops, ICDMW 2013 pp 162–171

[6] Kou Y, Lu CT, Sirwongwattana S, et al (2004) Survey of fraud detection techniques. Conference Proceeding - IEEE International Conference on Networking, Sensing and Control 2:749–754

[7] Kumar MS, Soundarya V, Kavitha S, et al (2019) Credit Card Fraud Detection Using Random Forest Algorithm. 2019 Proceedings of the 3rd International Conference on Computing and Communications Technologies, ICCCT 2019 pp 149–153

[8] Niveditha G, Abarna K, Akshaya GV (2019) Credit Card Fraud Detection Using Random Forest Algorithm. International Journal of Scientific Research in Computer Science, Engineering and Information Technology 5(2):301–306

[9] Puh M, Brki´c L (2019) Detecting credit card fraud using selected machine learning algorithms. 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics pp 1250–1255

[10] Ratih ID, Retnaningsih SM, Islahulhaq I, et al (2022) Synthetic minority over-sampling technique nominal continuous logistic regression for imbalanced data. AIP Conf Proc 2668(1):70021–70021