# Anomaly Detection in Corporate Credit Card Transactions using Machine Learning

## 1. Abstract

This project aims to develop a machine learning model to detect specific anomalous transaction patterns, such as duplicate payments (e.g., identical transactions processed multiple times), fraudulent activities (e.g., unauthorized or suspicious account activity), and violations of Delegation of Authority (DOA) policies (e.g., approvals exceeding role-based limits) for a publicly-traded company. The model will be trained on a diverse, synthetic dataset simulating real-world transactional scenarios to ensure privacy and accuracy. An automated documentation pipeline will be implemented to generate comprehensive audit documentation to meet Sarbanes-Oxley (SOX) compliance requirements, while reducing toil for users of the model, the company’s internal auditors. The model will be coded in Python, stored on GitHub, and deployed using Amazon SageMaker. Performance evaluation will focus on metrics such as accuracy, precision, and recall, with ongoing monitoring to adapt to evolving transaction patterns.

## 2. Introduction

The rapid growth of the commerce industry has led to an alarming increase in digital fraud and associated losses [1]. Detecting anomalous transaction patterns, such as duplicate payments, fraudulent payments, and transactions that do not follow the company’s Delegation of Authority (DOA), is crucial for maintaining a healthy financials of an organization [2].

## 3. Background and Motivation

The IT Audit team of a publicly traded company aims to enhance its operational efficiency by adopting automated data analytics techniques. Currently, significant time and resources are dedicated to the manual review and investigation of potential anomalies in corporate credit card transactions, such as duplicate payments and unauthorized expenditures. By developing a machine learning model to identify anomalous transaction patterns, the team seeks to streamline this process, allowing auditors to shift their focus to higher-value tasks, such as risk mitigation and strategic decision-making [3], [4]. This initiative aligns with the broader industry trend of leveraging advanced analytics to improve audit accuracy and reduce operational overhead.

## 4. Objectives

The objectives of the proposed project are as follows: 1. Develop a machine learning model that accurately detects anomalous transaction patterns 2. Create an automated documentation process that meets SOX reporting standards 3. Provide insights into the most significant variables and relationships in the data 4. Deploy the model on Amazon SageMaker for integration into the IT Audit team’s workflow

## 5. System Architecture

The proposed system architecture comprises four main components: data preprocessing, model development, model deployment, and documentation module. The data preprocessing component will handle data cleaning, anonymization, and feature engineering. The model development component will evaluate various machine learning algorithms, and implement most optimal. The model deployment component alongside documentation module will focus on integrating the trained model into the IT Audit team’s workflow using Amazon SageMaker.

## 6. Methodology

The methodology for this project involves several key stages, beginning with data preparation [5], [6]. The corporate credit card transaction dataset will be accessed and anonymized to protect sensitive information, using Python libraries such as Faker and DataSynthesizer. Preprocessing steps include data cleaning and transformation with Pandas and NumPy, followed by feature engineering techniques like scaling and encoding categorical variables to ensure the data is ready for analysis. The next phase, exploratory data analysis (EDA), will involve visualizing the dataset and computing descriptive statistics [7]. Tools like Matplotlib will be used to create informative plots and gain insights into data patterns and potential anomalies. This analysis will guide the selection of features and models for the subsequent development stage.In the model development phase, algorithms will be implemented using Scikit-learn and TensorFlow [8], [9], [10]. Techniques such as feature selection and cross-validation will be employed to optimize model performance, while challenges like class imbalance and high-dimensional data will be addressed to enhance robustness. For model evaluation, performance will be assessed using metrics like accuracy, precision, recall, F1-score, and ROC curves, provided by Scikit-learn [11]. The model’s generalization will also be examined to ensure a balance between bias and variance. To meet documentation and compliance requirements, an automated documentation pipeline will be developed using Python tools like pdoc3. This system will adhere to SOX reporting standards by incorporating specific guidelines and ensuring clear, consistent documentation. The project will conclude with deployment [12] using Amazon SageMaker, enabling practical application and scalability for corporate use.

## 7. Timeline

The project will be completed by December 8, 2024, to align with course requirements.

## 8. Expected Results

* A machine learning model that effectively detects anomalous transaction patterns
* An automated documentation process that meets SOX reporting standards
* Insights into feature importance and relationships in the data
* A deployed model on Amazon SageMaker, ready for integration into the IT Audit team’s workflow

## 9. Future Work

Following successful completion of this project, its results could be further iterated on to provide more value: - Exploring advanced feature engineering techniques and deep learning algorithms for improved performance [13], [14] - Adapting the automated documentation process to remove toil in documentation of other busienss processes

## 10. Conclusion

This project proposal presents a comprehensive approach to detecting anomalous transaction patterns in corporate credit card data using machine learning techniques. By leveraging state-of-the-art algorithms, automated documentation processes, and cloud-based deployment, the proposed solution aims to enhance the efficiency and effectiveness of the IT Audit team’s assurance efforts.

## References

[1] A. Herreros-Martínez, R. Magdalena-Benedicto, J. Vila-Francés, A. J. Serrano-López, and S. Pérez-Díaz, “Applied machine learning to anomaly detection in enterprise purchase processes,” arXiv:2405.14754 [cs], May 2023, doi: 10.48550/arXiv.2405.14754.

[2] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, “Enhancing credit card fraud detection: An ensemble machine learning approach,” Big Data Cogn. Comput., vol. 8, no. 1, p. 6, Jan. 2024, doi: 10.3390/bdcc8010006.

[3] A. Ali et al., “Financial fraud detection based on machine learning: A systematic literature review,” Appl. Sci., vol. 12, no. 19, p. 9637, Jan. 2022, doi: 10.3390/app12199637.

[4] A. Mutemi and F. Bacao, “E-commerce fraud detection based on machine learning techniques: Systematic literature review,” Big Data Min. Anal., vol. 7, no. 2, pp. 419–444, Jun. 2023, doi: 10.26599/BDMA.2023.9020023.

[5] A. Teymouri, M. Komeili, E. Velazquez, O. Baysal, and M. Genkin, “DATA5000OMBA – Accessing Data,” Carleton University BrightSpace, 2022. [Online]. Available: [https://brightspace.carleton.ca/d2l/le/content/288511/viewContent/3966339/View].

[6] A. Teymouri, M. Komeili, E. Velazquez, O. Baysal, and M. Genkin, “DATA5000OMBA – Pre-Processing Data,” Carleton University BrightSpace, 2022. [Online]. Available: [https://brightspace.carleton.ca/d2l/le/content/288511/viewContent/3966371/View].

[7] A. Teymouri, M. Komeili, E. Velazquez, O. Baysal, and M. Genkin, “DATA5000OMBA – Visualizing Data,” Carleton University BrightSpace, 2022. [Online]. Available: [https://brightspace.carleton.ca/d2l/le/content/288511/viewContent/3966372/View].

[8] A. Teymouri, M. Komeili, E. Velazquez, O. Baysal, and M. Genkin, “DATA5000OMBA – Machine Learning Introduction,” Carleton University BrightSpace, 2022. [Online]. Available: [https://brightspace.carleton.ca/d2l/le/content/288511/viewContent/3966341/View].

[9] A. Teymouri, M. Komeili, E. Velazquez, O. Baysal, and M. Genkin, “DATA5000OMBA – Machine Learning Algorithms - 1,” Carleton University BrightSpace, 2022. [Online]. Available: [https://brightspace.carleton.ca/d2l/le/content/288511/viewContent/3966344/View].

[10] A. Teymouri, M. Komeili, E. Velazquez, O. Baysal, and M. Genkin, “DATA5000OMBA – Machine Learning Algorithms - 2,” Carleton University BrightSpace, 2022. [Online]. Available: [https://brightspace.carleton.ca/d2l/le/content/288511/viewContent/3966347/View].

[11] A. Teymouri, M. Komeili, E. Velazquez, O. Baysal, and M. Genkin, “DATA5000OMBA – Evaluating Machine Learning Models,” Carleton University BrightSpace, 2022. [Online]. Available: [https://brightspace.carleton.ca/d2l/le/content/288511/viewContent/3966376/View].

[12] A. Teymouri, M. Komeili, E. Velazquez, O. Baysal, and M. Genkin, “DATA5000OMBA – Analytics Accelerators,” Carleton University BrightSpace, 2022. [Online]. Available: [https://brightspace.carleton.ca/d2l/le/content/288511/viewContent/3966370/View].

[13] A. Y. Shdefat et al., “Comparative analysis of machine learning models in online payment fraud prediction,” in 2024 Intelligent Methods, Systems, and Applications (IMSA), 2024, pp. 243–250. doi: 10.1109/IMSA61967.2024.10652861.

[14] A. Teymouri, M. Komeili, E. Velazquez, O. Baysal, and M. Genkin, “DATA5000OMBA – Role of the Data Scientist,” Carleton University BrightSpace, 2022. [Online]. Available: [https://brightspace.carleton.ca/d2l/le/content/288511/viewContent/3966369/View].