Optimizing Credit Card Fraud Detection Through Imbalanced Dataset Management and Feature Analysis

Ankitha Dongerkerry Pai
School of Computing and Augmented Intelligence
Arizona State University
apai14@asu.edu

Karthik Mahalingam
School of Computing and Augmented Intelligence
Arizona State University
kmahali2@asu.edu

Abstract

Our objective is to delve into strategies for enhancing credit card fraud detection systems, particularly addressing the challenges posed by imbalanced transaction data. In light of the significant threat that credit card fraud poses to businesses and consumers alike, our motivation stems from the imperative need to develop robust detection mechanisms that safeguard financial assets and uphold trust in digital transactions. By studying this topic, we aim to offer practical solutions to businesses, enabling them to better navigate the complexities of fraud detection in the modern landscape. Our analysis of transaction features such as time and amount seeks to uncover patterns associated with fraudulent activities, refining detection algorithms for heightened accuracy. Ultimately, our contributions aim to empower businesses to fortify their defenses against fraud, minimize disruptions from false alarms, and foster a safer digital economy for all stakeholders.

1 Motivation

Our project is motivated by the urgent need to tackle the growing menace of credit card fraud in today's digital landscape. With businesses and consumers increasingly reliant on electronic transactions, the risk of fraudulent activities poses a significant threat to financial security and trust. Through our research, we aim to provide practical solutions for businesses to strengthen their fraud detection mechanisms. By analyzing transaction data and uncovering patterns indicative of fraud, we seek to enhance the accuracy of detection algorithms. Ultimately, our efforts aim to empower businesses to fortify their defenses against fraud, ensuring a safer and more secure digital economy for all stakeholders.

2 Current Progress

2.1 Achievements

We are going on track as per our proposed timeline mentioned in the project proposal. We have laid the initial groundwork for our project by preparing our computational environment and importing the dataset for analysis. The preliminary steps we have taken include:

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

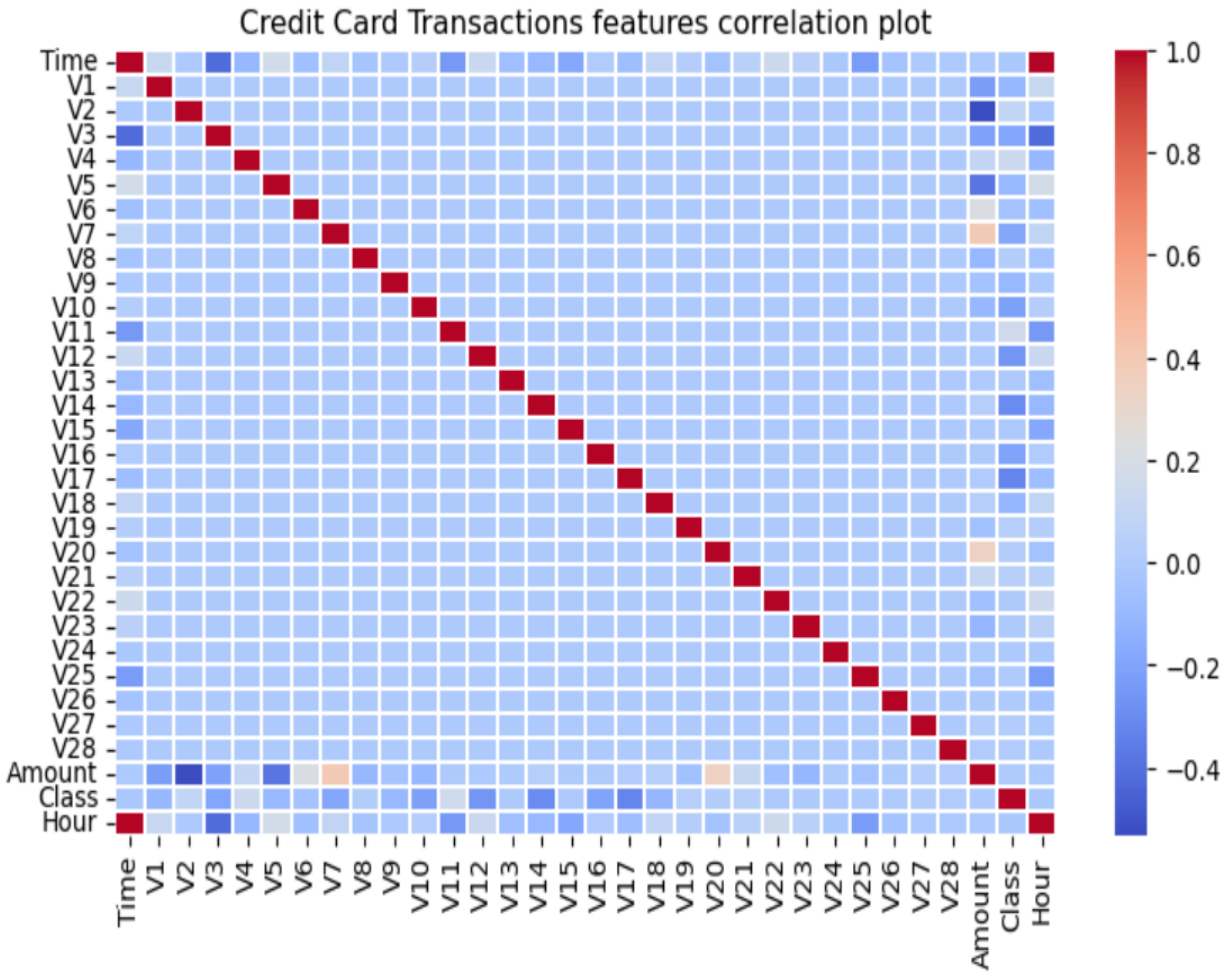

Figure 1: Correlation matrix heatmap of credit card transaction features, displaying the degree of linear relationship between variables.

- Importing necessary Python libraries for data manipulation and visualization.
- Reading the credit card transactions dataset into a pandas DataFrame.
- Beginning our exploratory data analysis (EDA) to better understand the dataset's structure and the nature of its contents through basic statistics and visualizations.
- Through the visualization we understood the importance of features via co-relation matrix.
- Anomaly detection was done on the data visualizations. We performed Random forest classifier and got an accuracy of 85%.

2.2 Difficulties

Fraud detection is a very highly popular choice in machine learning and data science, and there were a lot of suggestions on which models work the best. It was overwhelming to choose a suitable model. Then, through more research, we found that doing EDA would help us figure out more about our dataset, helping us make a better choice of model

3 Future Plan

Moving forward, our plans include:

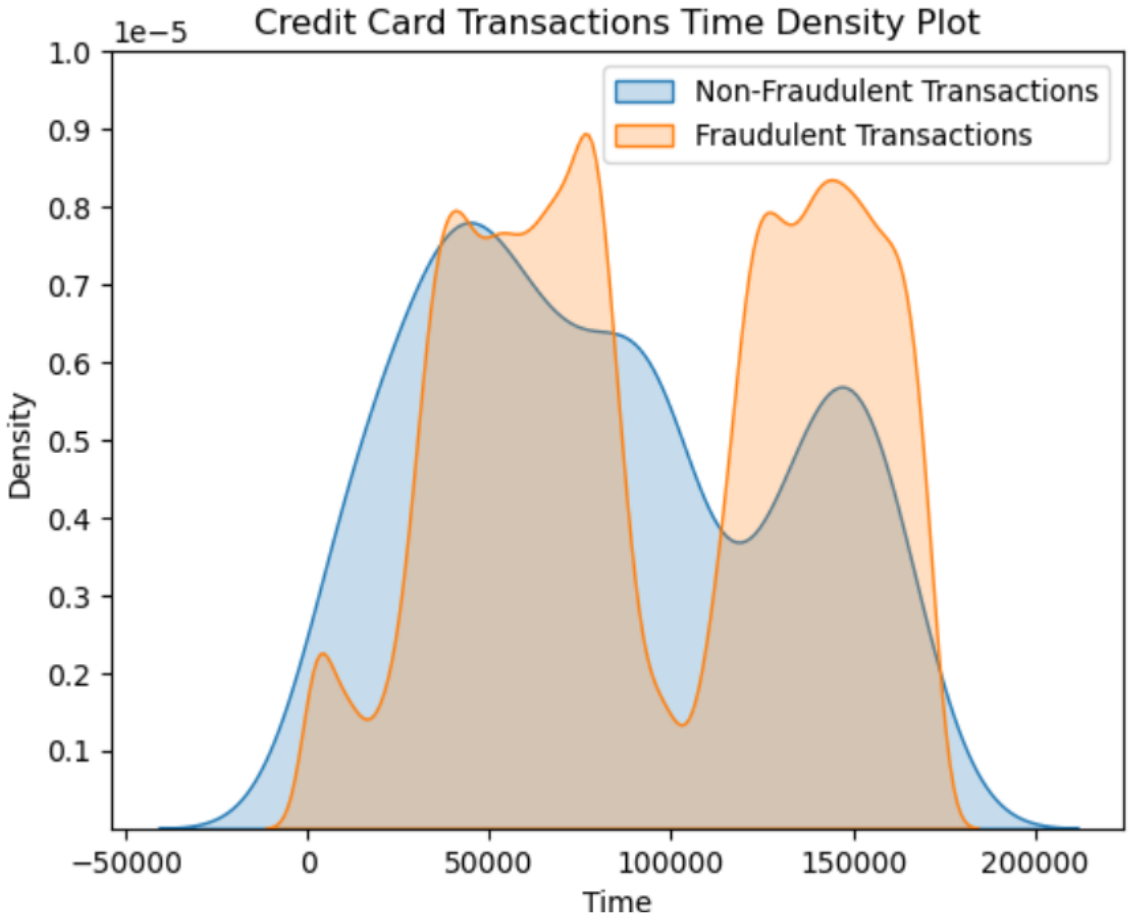

Figure 2: Density plot comparing the distribution of transaction times for non-fraudulent and fraudulent credit card transactions.

- Hyperparameter tuning will be conducted to optimize model performance, leading to better accuracy.
- Testing various machine learning models to identify the most effective technique for detecting fraudulent transactions, which could range from logistic regression and decision trees to random forests and neural networks.
- Complete the final project report.

3.1 Unexpected Discoveries

As we continue with our project, we're ready for any surprises we might encounter in our data or model performance. If we find that certain features aren't helping as much as we thought they would, we'll take a closer look at how we're selecting or changing those features.

References

[1] Kaggle. (2024). Credit Card Fraud Detection. Kaggle. Retrieved from https://drive.google.com/file/d/1JXS4vbOXuBbYEB-cUiY-hrDMGGgfbzRG/view?usp=sharing

[2] R. Wang and G. Liu, "Ensemble Method for Credit Card Fraud Detection," 2021 4th International Conference on Intelligent Autonomous Systems (ICoIAS), Wuhan, China, 2021, pp. 246-252, doi: 10.1109/ICoIAS53694.2021.00051.

[3] B. A. Smadi, A. A. S. AlQahtani and H. Alamleh, "Secure and Fraud Proof Online Payment System for Credit Cards," in “2021 IEEE 12th Annual Ubiquitous Computing, Electronics Communication Conference (UEMCON),” New York, NY, USA, 2021, pp. 0264-0268, doi: 10.1109/UEMCON53757.2021.9666549.

[4] Marazqah Btoush EAL, Zhou X, Gururajan R, Chan KC, Genrich R, Sankaran P. “A systematic review of literature on credit card cyber fraud detection using machine and deep learning.” PeerJ Comput Sci. 2023 Apr 17;9:e1278. doi: 10.7717/peerj-cs.1278. PMID: 37346569; PMCID: PMC10280638.

[5] A. Maurya and A. Kumar, "Credit card fraud detection system using machine learning technique," in “2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom),” Malang, Indonesia, 2022, pp. 500-504, doi: 10.1109/CyberneticsCom55287.2022.9865466.

[6] I. Vejalla, S. P. Battula, K. Kalluri and H. K. Kalluri, "Credit Card Fraud Detection Using Machine Learning Techniques," in “2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS),” Nagpur, India, 2023, pp. 1-4, doi: 10.1109/PCEMS58491.2023.10136040.

[7] https://www.kaggle.com/code/gpreda/credit-card-fraud-detection-predictive-models