Project Report: Enhancing Credit Card Fraud Detection with Generative AI

Author – Ayush Choudhary

Institute – INDIAN INSTITUTEE OF TECHNOLOGY, ROORKEE

1. Abstract

Credit card fraud is a significant challenge in the financial industry, exacerbated by the severe class imbalance in transaction datasets, where fraudulent cases are extremely rare. This project addresses the issue by employing a Generative Adversarial Network (GAN) to synthesize realistic fraudulent transaction data. The synthetic data was used to augment the original dataset, creating a more balanced training environment for a machine learning classifier. By comparing a model trained on the original imbalanced data with one trained on the augmented data, this report demonstrates that the Generative AI approach significantly improves the model's ability to detect fraud, particularly enhancing the **recall** and **F1-score** for the minority class.

2. Introduction

2.1. Problem Statement

Standard machine learning models often struggle with imbalanced datasets. In the context of fraud detection, the model can achieve high accuracy by simply predicting "not fraud" for every transaction, yet fail to identify the rare fraudulent cases that are of primary interest. The core challenge is to improve the detection of the minority (fraud) class without generating an unmanageable number of false alarms (false positives).

2.2. Proposed Solution

This project leverages a **Conditional Tabular Generative Adversarial Network (CTGAN)** to learn the underlying data distribution of fraudulent transactions. By training the GAN exclusively on these rare instances, we can generate a high volume of new, synthetic data points that are statistically similar to real fraud. This synthetic data is then used to augment the original dataset, creating a balanced training set for a **Random Forest classifier/ XGboost classifier**.

3. Methodology

3.1. Dataset

The project utilized the "Credit Card Fraud Detection" dataset from Kaggle.

- **Features:** 30 numerical features (`Time`, `Amount`, and 28 anonymized PCA components `V1-V28`).
- **Target:** `Class` (0 for Normal, 1 for Fraud).
- **Imbalance:** Fraudulent transactions constitute only **0.17%** of the dataset, presenting a significant modeling challenge.

3.2. Exploratory Data Analysis (EDA)

Initial analysis confirmed the severe class imbalance. The distributions of `Time` and `Amount` were also analyzed to understand the dataset's characteristics before modeling.

3.3. Synthetic Data Generation with CTGAN

The CTGAN model from the **Synthetic Data Vault (SDV)** library was trained exclusively on the real fraud samples (`Class == 1`). This focused approach ensures the generator becomes an expert at creating realistic fraud data. After training, the model generated new synthetic fraud samples to be added to our training data.

3.4. Model Training

Two XGBoost classifier were trained for comparison:

1. **Baseline Model:** Trained on the original, severely imbalanced dataset.
2. **Augmented Model:** Trained on the augmented dataset, which combines the original data with the newly generated synthetic fraud data.

Both models were evaluated on the same untouched test set, composed entirely of real data, to ensure a fair and realistic comparison.

4. Results and Discussion

4.1. Model Performance Comparison

The performance of the two models was evaluated using precision, recall, and F1-score for the fraud class.

Model	Precision (Fraud)	Recall (Fraud)	F1-Score (Fraud)
Baseline Model	0.91	0.76	0.83
Augmented Model	0.86	**0.93**	**0.89**

Export to Sheets

The results clearly show the benefit of data augmentation. While there was a slight decrease in precision, the **Augmented Model** achieved a **significant increase in recall from 0.76 to 0.89**. This means the new model successfully identified 89% of all actual frauds in the test set, a marked improvement over the baseline. The F-score, which balances precision and recall, indicating a better overall model(it tells us about how both parameters precision and recall are handelled, weather we have not sacrificed one because of other).

4.2. Visual Evaluation of Synthetic Data

To qualitatively assess the synthetic data, PCA were used to reduce the data's dimensionality and visualize the distributions of real vs. synthetic fraud samples.

5. Conclusion

This project successfully demonstrated the power of Generative AI in solving the critical problem of class imbalance in fraud detection. By augmenting a real-world dataset with high-quality synthetic data from a CTGAN, we were able to train a Random Forest model/(or can use any other classifier) that was significantly more effective at identifying fraudulent transactions. The key takeaway is that data augmentation with generative models is a powerful technique to build more robust and reliable machine learning systems, especially when dealing with rare events. This approach leads to tangible benefits, such as reduced financial losses and increased customer trust.