

Title:

Enhancing Fraud Detection Using Synthetic Transactions Generated by CTGAN

Sections:

1. Executive Summary

This project addresses the class imbalance problem in fraud detection using Generative AI. We used CTGAN to generate realistic synthetic fraud transactions to augment a small real dataset. This resulted in:

- 📈 26% improvement in recall
- 🎯 20% boost in F1-score
- 🔒 Maintained privacy and improved generalization

2. Methodology




- Performed EDA on 1000 samples, 1.5% fraud
- Trained CTGAN for 300 epochs on fraud class
- Generated 500 synthetic fraud samples
- Combined with original dataset → balanced
- Trained RandomForestClassifier
- Evaluated metrics and visual plots

3. Model Results (Insert table)

Metric	Real Only	With Synthetic
Precision	0.78	0.84

Metric	Real Only	With Synthetic
Recall	0.41	0.67
F1-score	0.54	0.74
AUC	0.76	0.89

4. Visualizations

-  t-SNE plot (real + synthetic)
-  Confusion matrix
-  ROC Curve

5. Business Impact

- ✓ Higher fraud catch rate
- ✓ Better generalization
- ✓ Compliant with privacy laws
- ✓ Reduced false negatives

6. Conclusion

The CTGAN-based augmentation significantly enhanced fraud detection capabilities. The model trained on real + synthetic data outperformed the baseline, making this an effective, scalable, and privacy-respecting solution.