PRIVACY-PRESERVING CREDIT CARD FRAUD DETECTION

Building a secure ML model to detect fraudulent transactions while protecting user privacy

Presented By : Rishank kumbhare (22070521184)
Divyanshu Dharmik(22070521174)
Akash Ujawane(22070521175)

DATASET OVERVIEW

We used a publicly available dataset from Kaggle with 284,807 transactions. It contains anonymized features (V1 to V28) transformed using PCA, along with Time, Amount, and Class labels. Only 0.17% of transactions are fraudulent, making it highly imbalanced.

Time	V1	V2	V3	...	V28	Amount	Class
0	-1.3598	-0.0728	2.5363	...	0.0218	149.62	0
1	1.1918	0.2661	0.1664	...	0.0003	2.69	0
2	-1.3584	-1.3402	1.7732	...	0.0102	378.66	0

PRIVACY-PRESERVING TECHNIQUES

- No personal details were used — no names, card numbers, or addresses.
- Data features were anonymized using PCA, hiding real identities.
- Model only saw transformed values, not actual customer info.
- Impossible to reverse the data back to original users.
- All processing was done securely to prevent data leaks.

METHODOLOGY

- Data Preprocessing: Cleaned and scaled features like Amount, handled class imbalance.
- Train-Test Split: Divided data to train the model and test its performance.
- Model Training: Used XGBoost — fast, accurate, and handles imbalanced data well.
- Privacy Preserved: Model trained only on anonymized features.

HANDLING IMBALANCED DATA

- Fraudulent transactions make up less than 0.2% of the data.
- Trained the model using class weighting to focus more on fraud cases.
- This prevents the model from always predicting "not fraud" just to boost accuracy.
- Goal: Improve recall (catch more frauds) without too many false alarms.

MODEL USED

- Tried models like Logistic Regression and Random Forest.
- XGBoost performed best — it's fast, accurate, and handles imbalanced data well.
- Automatically gives more weight to rare fraud cases.
- Easy to tune and works well with anonymized numeric data.

What is XGBoost?

- Builds a series of decision trees
- Each tree fixes errors made by the previous one
- Combines all trees for better accuracy

Why XGBoost?

- High precision and recall
- Robust against overfitting
- Scalable for large datasets

EVALUATION METRICS

- Accuracy – Overall correctness of predictions
- Precision – Out of all predicted frauds, how many were actually fraud
- Recall – Out of all actual frauds, how many we caught
- F1-Score – Balance between precision and recall
- ROC-AUC – Measures model's ability to distinguish fraud vs non-fraud

RESULTS & MODEL COMPARISON

Key Takeaway:

- **XGBoost outperformed other models in all key metrics.**
- **Best balance between catching frauds (recall) and being accurate (precision).**

Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	97.5	66	62	64
Random Forest	99.3	89	85	87
XGBoost	99.6	92	90	91

FUTURE ENHANCEMENTS

- Add real-time detection to catch fraud as it happens.
- Integrate Differential Privacy to mathematically guarantee user data protection.
- Use Federated Learning to train models across devices without sharing data.
- Automate retraining with fresh data to adapt to new fraud patterns.
- Deploy as a secure API for easy integration with banking systems.

CONCLUSION

- We built a smart model that spots credit card fraud.
- It works without using any personal or sensitive info.
- XGBoost gave the best results with great accuracy and speed.
- The model handles rare fraud cases really well.

THANK YOU