FRAUD DETECTION IN FINANCIAL TRANSACTIONS

Mani Jose | Nakul Krishna R

School of Mathematics and Statistics, University College Dublin

ABSTRACT

This project develops a real-time fraud detection system using the synthetic BankSim dataset, simulating realistic financial transactions with rare fraud cases (~1.2%). We apply advanced feature engineering and train a **LightGBM ensemble with undersampling** and Optuna hyperparameter tuning, achieving **AUROC 0.9993** and **AUPRC 0.9578**. A FastAPI service with an Synthetic Data Vault (SDV) streaming -powered streaming dashboard enables low-latency scoring and continuous evaluation in a production-like setup.

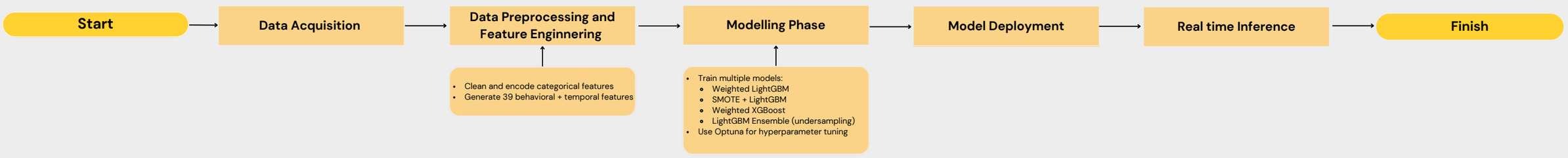
DATASET

The **BankSim** dataset, a synthetic yet behaviorally realistic collection of financial transactions for fraud detection research is used for this project.

- **Transactions: 594,643 records**
- **Fraud Rate: ~1.2%** (reflecting real-world imbalance)
- **Features : 9 original features + Target Variable :** Fraud
  - 🆔 **Identifiers & Demographics** : customer, merchant, age, gender
  - ⏱️ **Transaction details** : step, category, amount
  - 🌍 **Location Data** : zipcodeOri, zipMerchant

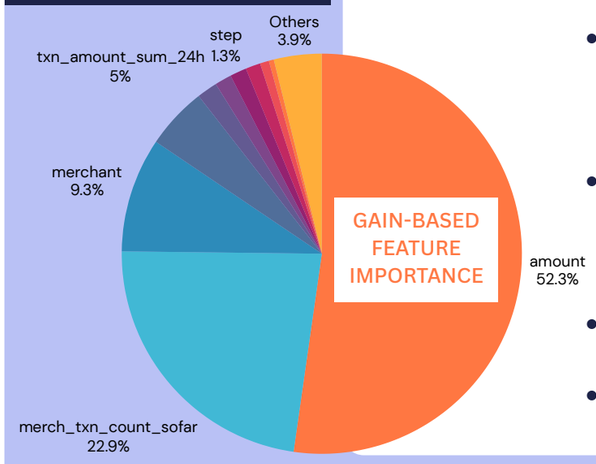
METHODOLOGY

- **Data Collection**: BankSim dataset — simulated banking transactions (timestamps, amounts, IDs, fraud labels).
- **Feature Engineering**: 39 **behavioral, temporal**, and **relational** features, ordered chronologically to avoid leakage.
- **Model Training**: Tested SMOTE, weighted loss, undersampling; **LightGBM ensemble (0.05 undersampling, Optuna-tuned)** achieved best performance.
- **Real-Time Inference**: **FastAPI scoring + SDV streaming** → live monitoring in the Streamlit dashboard.



FEATURE ENGINEERING

39 engineered features — designed to avoid data leakage and optimized for both batch and streaming inference.

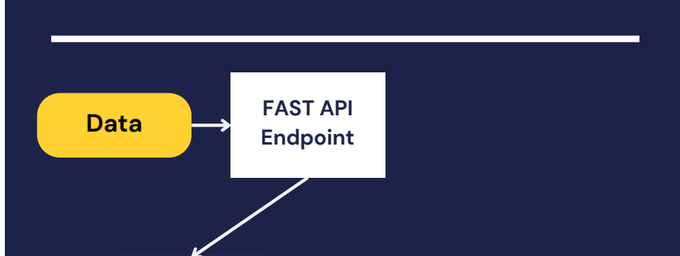


- **Temporal features** – amount_zscore_24h (deviation from 24h history), txn_count_24h / mean_amount_24h (velocity & average size), step_delta (time since previous transaction).
- **Customer–Merchant Interaction** – merchant_occurrences (past transactions), is_first_time_pair (new interaction), unique_merchants_count (merchant diversity).
- **Time-of-Day** – is_night flag for 00:00–06:00, a high-risk fraud window.
- **Categorical Encoding:** one-hot encoding for category, gender, age.

Gain-based feature importance from the final LightGBM ensemble showed that a small subset of these features provided most of the model's predictive power in both batch and streaming detection.

BEST MODEL AND DEPLOYMENT

- **Architecture:** LightGBM gradient boosting, ensemble of undersampled learners
- **Class Imbalance Handling:** Random undersampling (5% majority class kept) + ensemble averaging
- **Hyperparameter Tuning:** Optuna, 50+ trials, AUROC-optimized
- **Training Data Size:** 594k transactions (~1.2% fraud rate)
- **Deployment:** Batch + real-time scoring supported; integrated with Streamlit monitoring dashboard

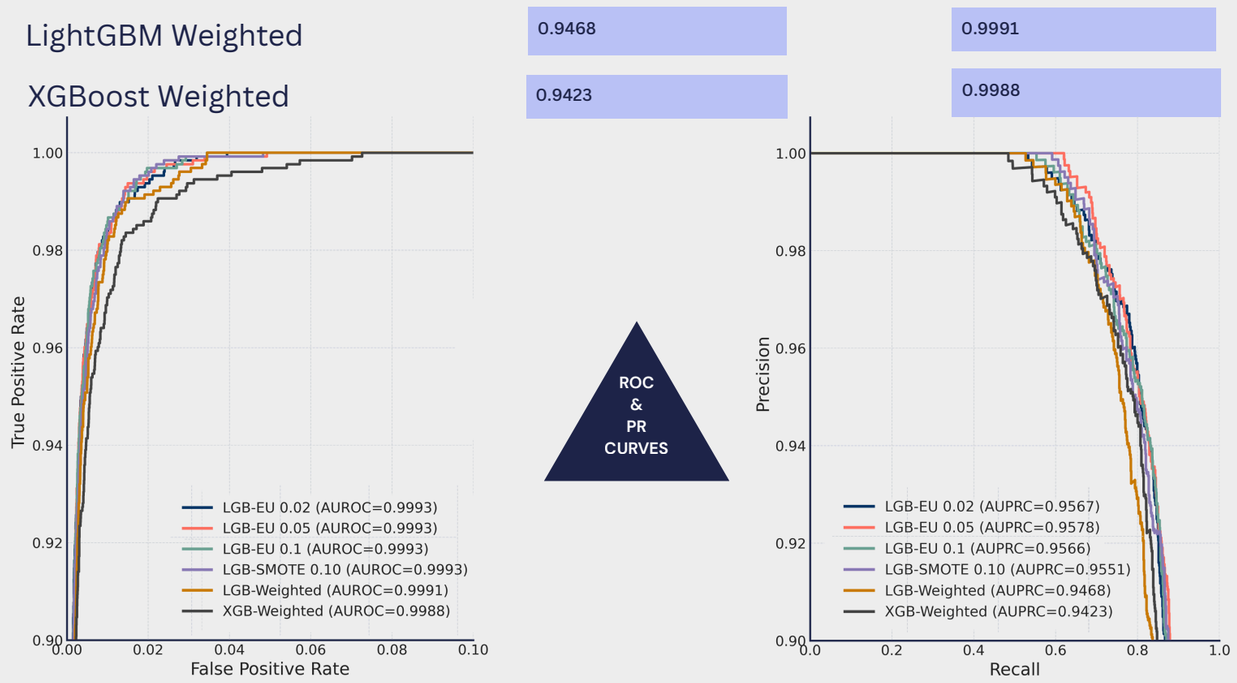


Confusion Matrix
High true-positive and true-negative counts with minimal misclassifications at a 0.5 decision threshold.

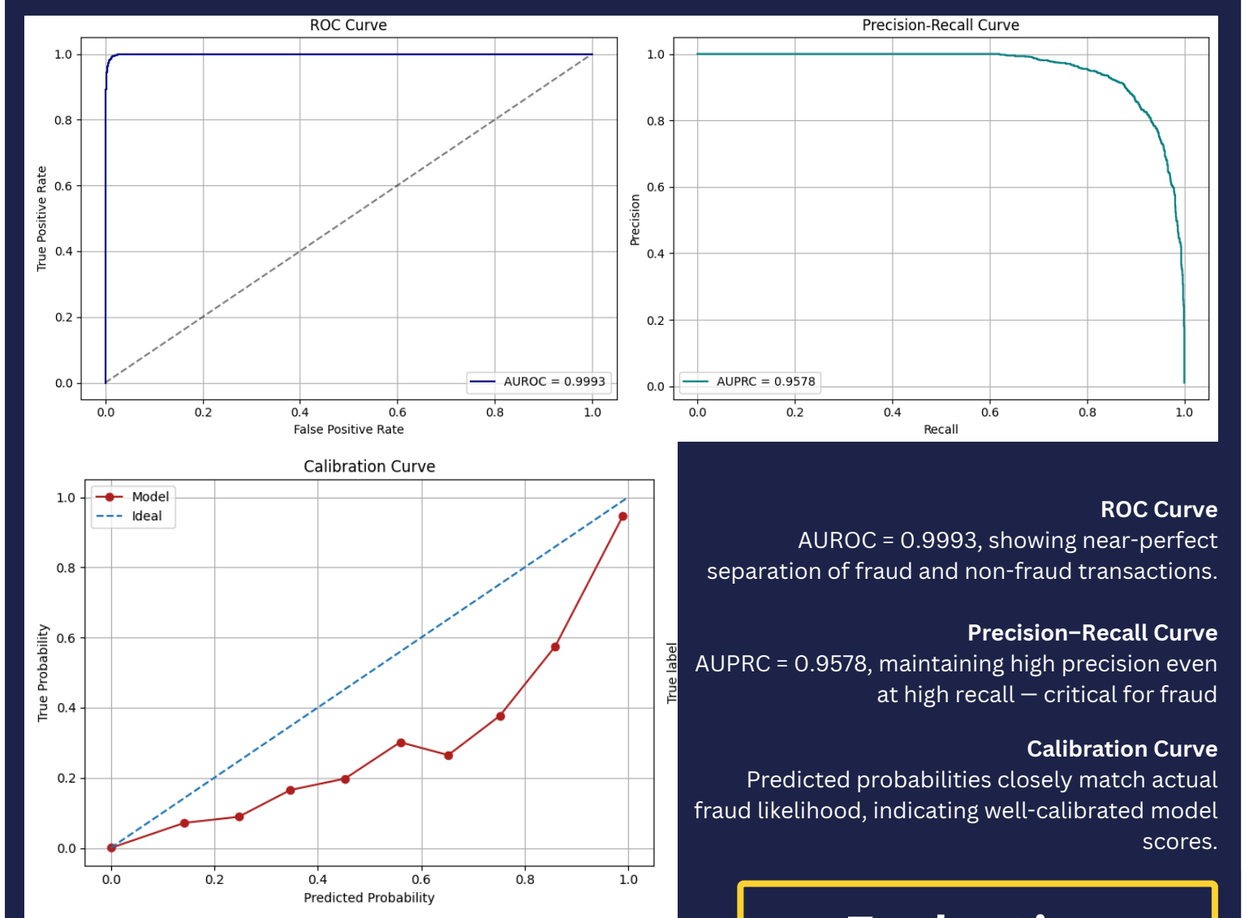
MODEL COMPARISON

Multiple models were evaluated to address extreme class imbalance and maximize fraud detection performance. Each model was tuned using Optuna, and evaluated using AUROC and AUPRC, which are robust to class imbalance.

MODEL	AUPRC	AUROC
LightGBM Ensemble 0.05	0.9578	0.9993
LightGBM Ensemble 0.02	0.9567	0.9993
LightGBM Ensemble 0.10	0.9566	0.9993
LightGBM + SMOTE 0.10	0.9551	0.9993
LightGBM Weighted	0.9468	0.9991
XGBoost Weighted	0.9423	0.9988



Receiver Operating Characteristic (ROC) and Precision–Recall (PR) curves for all evaluated models, the LightGBM Ensemble Model with 0.05 undersampling achieved the highest overall performance, with AUROC = 0.9993 and AUPRC = 0.9578. Both curves confirm strong discrimination ability and robustness under extreme class imbalance.



ROC Curve
AUROC = 0.9993, showing near-perfect separation of fraud and non-fraud transactions.

Precision–Recall Curve
AUPRC = 0.9578, maintaining high precision even at high recall — critical for fraud

Calibration Curve
Predicted probabilities closely match actual fraud likelihood, indicating well-calibrated model scores.

Evaluation Metrics

AUROC - 0.9993
AUPRC - 0.9578
Latency - <10 ms

FUTURE SCOPE

- **Graph modeling:** Detect fraud rings via GNNs on customer–merchant networks.
- **Online learning:** Adapt models to evolving fraud patterns.
- **Feature enrichment:** Integrate device, geo, and behavioral signals.
- **Interpretability:** Add SHAP/LIME for real-time explainability.

REFERENCES

- López-Rojas & Axelsson (2014) – BankSim: A bank payment simulation for fraud detection research.
- Chawla et al. (2002) – SMOTE: Synthetic Minority Over-sampling Technique.
- Chen & Guestrin (2016) – XGBoost: A scalable tree boosting system.
- Ke et al. (2017) – LightGBM: A highly efficient gradient boosting decision tree.

manijosevg@gmail.com | nakulkrishna96@gmail.com