Hackathon edition#3

Credit Card Fraud Detection:
Model Comparison and Evaluation

Team Details :

Manish Tiwari (Leader) (IITMCS_2406327)
Anushka Sarkar (IITMCS_2406255)
Jay Paneliya (IITMCS_24091607)

Project link
https://github.com/JAYpaneliya/credit-card-fraud-detection

Problem Statement

In real-world applications, fraud detection remains one of the most critical and complex classification challenges due to highly imbalanced datasets and evolving fraud patterns. The objective of this project was to:

- Train and evaluate multiple machine learning classification models
- Compare their performance on key metrics (Precision, Recall, F1-score, etc.)
- Perform hyperparameter tuning to improve model effectiveness
- Handle class imbalance to ensure realistic fraud detection capabilities
- Build an interactive dashboard for comparison, visualization, and exploration

Dataset Overview

- Source: Kaggle (European cardholder data)
- 284,807 transactions
- 492 fraud cases
- Features: PCA-anonymized (V1 to V28), Time, Amount
- Target: Class (0 = non-fraud, 1 = fraud)

Data Preprocessing

- Scaled 'Time' and 'Amount' using StandardScaler
- Applied SMOTE on training set
- 80/20 train-test split
- Test set kept imbalanced for realistic evaluation

Models used

- Logistic Regression
- Random Forest
- Support Vector Machine
- Gradient Boosting
- XGBoost

Each trained on SMOTE-balanced data

Evaluation metrics

- Precision: Correctly predicted frauds
- Recall: Caught fraud out of all frauds
- F1-score: Balance between precision and recall
- ROC-AUC: Overall classification strength

Hyperparameter tuning

- Applied on XGBoost using RandomizedSearchCV
- Optimized F1-score
- Tuned parameters: learning rate, depth, estimators
- Result: Improved recall and F1 performance

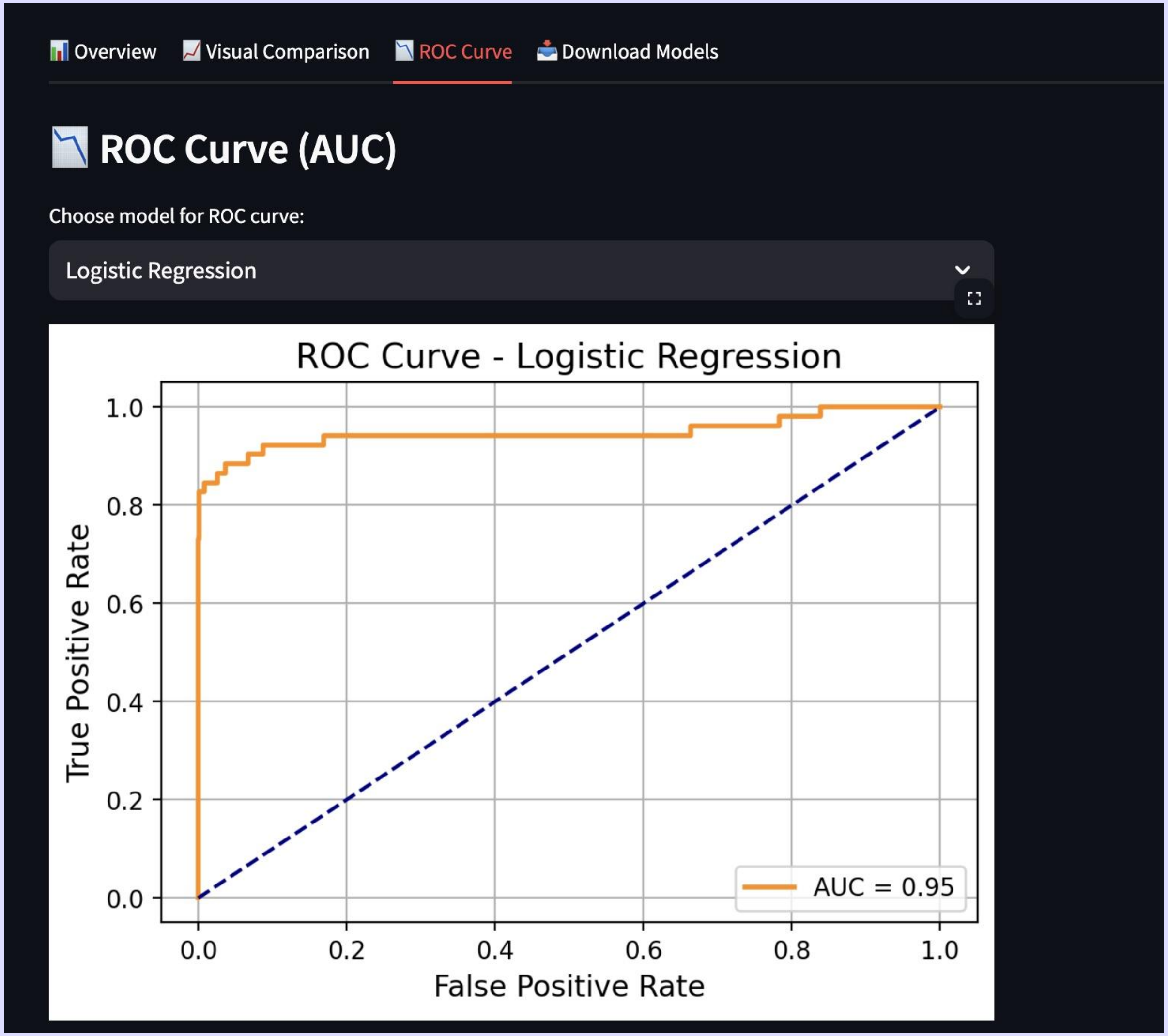
Streamlit Dashboard features

- Overview Tab: metric table for all models
- Visual Comparison: metric-based bar plots + confusion matrix
- ROC Curve: AUC plots per model
- Download Tab: trained .pkl models available

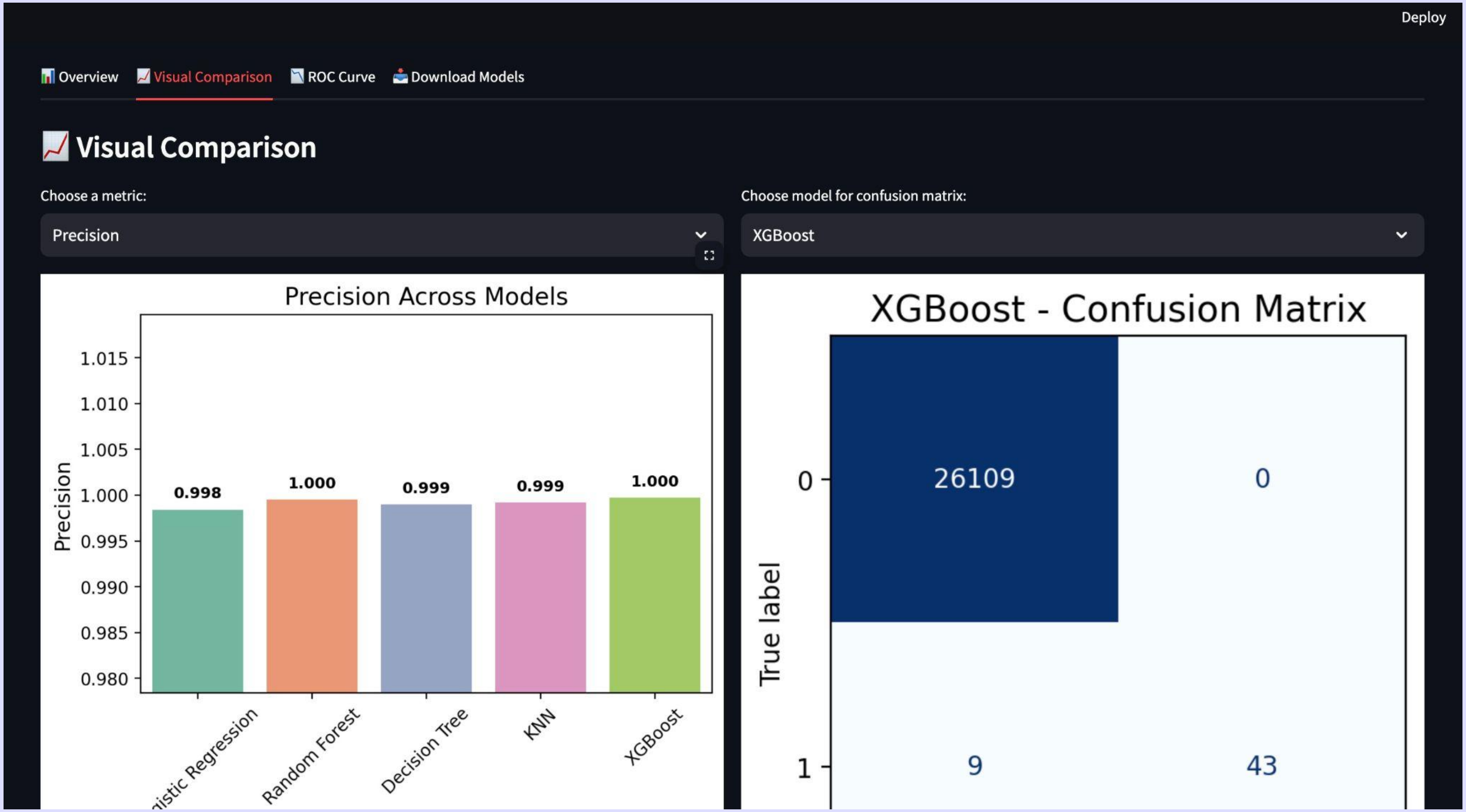
Streamlit Dashboard features

	Model	Accuracy	Precision	Recall	F1-score
0	Logistic Regression	0.998600	0.998400	0.998600	0.998500
1	Random Forest	0.999500	0.999500	0.999500	0.999500
2	Decision Tree	0.999000	0.999000	0.999000	0.999000
3	KNN	0.999200	0.999200	0.999200	0.999200
4	XGBoost	0.999700	0.999700	0.999700	0.999600

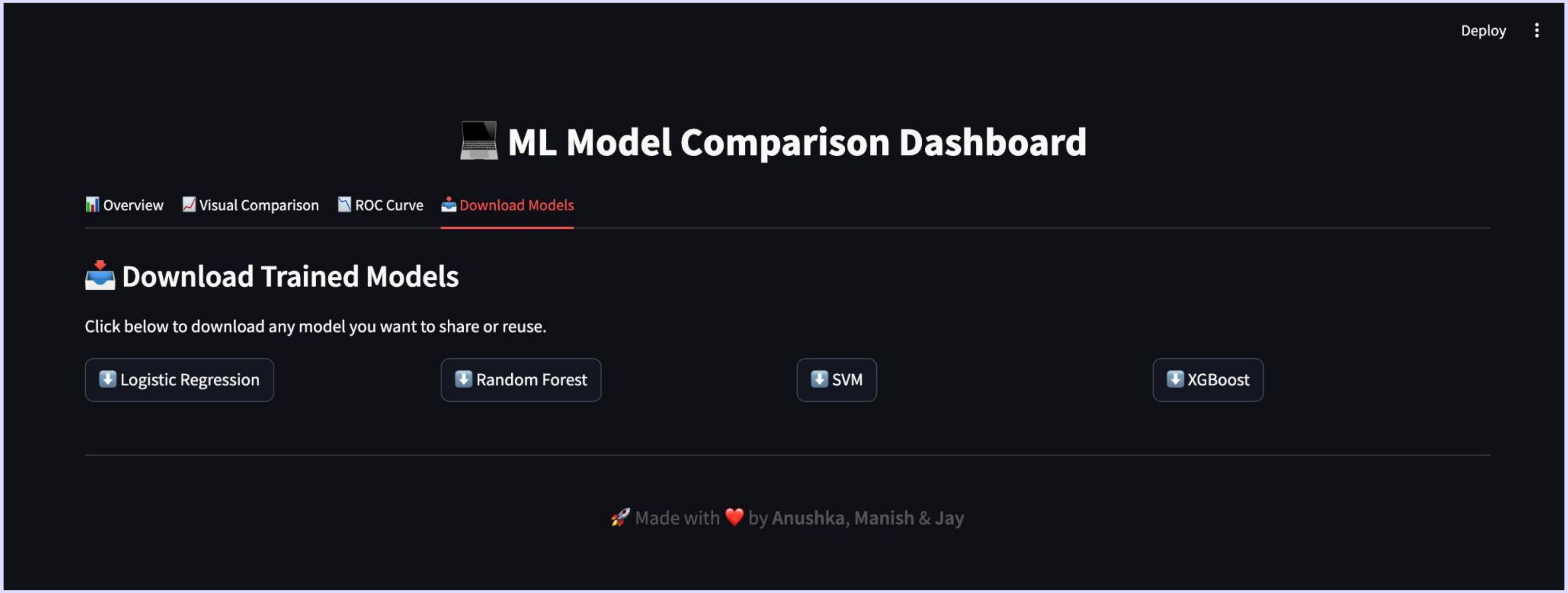
Streamlit Dashboard features



Streamlit Dashboard features



Streamlit Dashboard features



Key Insights

- XGBoost (tuned) performed best
- SMOTE drastically improved recall
- Trade-off: SVM/LogReg high precision, low recall
- Visualizations made comparisons easier
- SHAP helped interpret feature influence

THANK YOU

Presented by

Jay, Manish & Anushka