

Credit Card Fraud Detection Project

Credit Card Fraud Detection

Detecting fraudulent credit card transactions using Machine Learning.

Dataset

- Source: Kaggle - Credit Card Fraud Detection
- Total Transactions: 284,807
- Fraudulent Transactions: 492 (Severely imbalanced)
- Features:
- Time, Amount (scaled)
- V1 to V28: PCA-transformed anonymized features
- Class: Target variable (0 = Legit, 1 = Fraud)

Tools & Libraries Used

- Platform: Python, Jupyter Notebook
- Libraries:
- Data Handling: pandas, numpy
- Visualization: matplotlib, seaborn
- ML Models: scikit-learn, xgboost

Project Workflow

1. Data Loading & Exploration

- Checked shape, null values
- Verified severe class imbalance (fraud is ~0.17%)

Credit Card Fraud Detection Project

- Basic visualizations of Amount and Time distributions

2. Preprocessing

- Feature Scaling: Scaled Time and Amount using StandardScaler
- Train-Test Split:
 - 80% train, 20% test
 - Used StratifiedShuffleSplit to maintain fraud ratio

3. Model Training

Trained the following models:

- Logistic Regression
- Random Forest Classifier
- XGBoost Classifier

4. Model Evaluation

Model	Accuracy	Precision	Recall	F1 Score
-------	----------	-----------	--------	----------

-----	-----	-----	-----	-----
-------	-------	-------	-------	-------

Logistic Regression	~97.5%	Moderate	Lower	Fair
---------------------	--------	----------	-------	------

Random Forest	~99.3%	High	High	Very Good
---------------	--------	------	------	-----------

XGBoost	~99.5%	Very High	Very High	Excellent
---------	--------	-----------	-----------	-----------

- Confusion matrices and classification reports were used to assess performance.
- XGBoost was the best performer, achieving the highest recall with minimal false positives.

Conclusion

Credit Card Fraud Detection Project

- Imbalanced classification handled without SMOTE to avoid oversampling bias.
- XGBoost emerged as the most effective model.
- Focused on Recall to minimize false negatives (i.e., undetected frauds).

Future Enhancements

- Hyperparameter Tuning using GridSearchCV or RandomizedSearchCV
- Model Deployment with Flask or Streamlit
- Build a Dashboard for real-time fraud detection and analytics
- Explore advanced techniques for handling imbalanced datasets like:
 - Ensemble with anomaly detection
 - Cost-sensitive learning

File Structure

credit-card-fraud-detection/

creditcard.csv

fraud_detection.ipynb

requirements.txt

README.md

models/

logistic_model.pkl

rf_model.pkl

xgb_model.pkl

Key Learnings

- Real-world datasets are often imbalanced metrics like Recall matter more than accuracy.

Credit Card Fraud Detection Project

- Tree-based models like Random Forest and XGBoost handle such data well out-of-the-box.
- Explainability and false positive control are crucial in fraud detection tasks.