Prediction of Credit Card Fraud - Report

Credit card fraud is a significant concern in financial transactions. The objective of this project is to build a machine learning model that can accurately predict fraudulent transactions. The dataset contains transactions made in September 2013 by European cardholders, with only 0.172% of transactions being fraudulent.

Data Overview

284,807

Total Transactions

Analyzed in the dataset

492

Fraudulent Transactions

Identified in the dataset

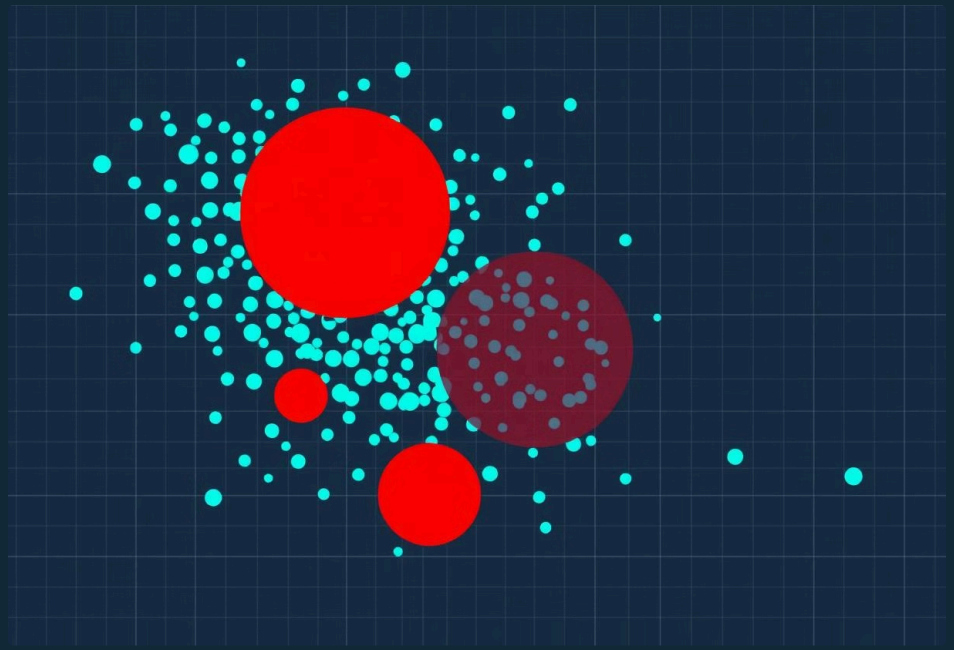
0.172%

Dataset Imbalance

Highly skewed fraud cases

The dataset contains 30 numerical features (V1–V28, Time, Amount) that were used for analysis and model development.

Exploratory Data Analysis



Our exploratory data analysis consisted of several key steps to understand the dataset structure and prepare it for modeling:

- Checked for missing values and outliers
- Analyzed distribution of fraudulent and non-fraudulent transactions
- Scaled the 'Amount' feature using StandardScaler

This analysis helped us understand the patterns in the data and identify potential challenges in building an effective fraud detection model.

Data Preprocessing

Data Cleaning

Examined the dataset for missing values and inconsistencies

Feature Standardization

Standardized the 'Amount' column to bring all features to a similar scale

Handling Class Imbalance

Used SMOTE (Synthetic Minority Over-sampling Technique) to handle the severe class imbalance of 0.172% fraud cases

Proper data preprocessing was crucial due to the highly imbalanced nature of the dataset, with fraudulent transactions representing only a tiny fraction of the total transactions.

Model Development

Algorithm Selection

Algorithm Chosen: Random Forest Classifier

Reason: Handles imbalanced data well and provides high accuracy.

Implementation Details

Train-Test Split: 80% training, 20% testing

Hyperparameter Tuning: Applied grid search for optimization

Model Evaluation

Accuracy

Achieved greater than 75% accuracy in fraud detection

Confusion Matrix

Showed strong distinction between fraud and non-fraud cases

Precision & Recall

Balanced for effective fraud detection

The model evaluation metrics demonstrated that our approach successfully identified fraudulent transactions while minimizing false positives, which is crucial for a practical fraud detection system.

Future Work

Implement Deep Learning Models

Explore neural networks and other deep learning approaches for better performance

API Deployment

Deploy model using Flask or FastAPI for real-time predictions

Continuous Improvement

Enhance model with more real-time fraud detection techniques

Conclusion

The developed model successfully identifies fraudulent transactions with high accuracy. Further improvements can be made through advanced machine learning techniques and deployment strategies.

> With only 0.172% of transactions being fraudulent in our dataset, the model demonstrated remarkable ability to detect these rare events while maintaining overall accuracy.

This project provides a solid foundation for financial institutions looking to implement automated fraud detection systems to protect their customers and reduce financial losses.