Credit Card Fraud Detection Using Machine Learning

Analyzing Transaction Patterns and Developing Predictive Models for Real-Time Fraud Detection

---

Author : Anushka Rajesh Salvi

Index:

Introduction:

This project aims to develop a machine learning-based system for detecting fraudulent credit card transactions using a dataset from Kaggle covering January 2019 to December 2020. By analyzing transaction features such as amount, merchant information, time of transaction, and customer demographics, the goal is to identify patterns in fraudulent behavior and build a model that predicts fraud with high accuracy. The insights from this analysis will help financial institutions enhance their fraud detection systems, minimize false positives, and reduce financial losses caused by fraudulent transactions.

Executive Summary:

This report presents actionable insights from the Credit Card Fraud Detection Analysis using machine learning models to predict fraudulent transactions. By analyzing historical data from January 2019 to December 2020, this project investigates fraud patterns based on transaction characteristics, merchant data, and timing to offer predictive solutions for better fraud detection in financial systems.

Key Insights:

1. Merchant-Level Fraud Analysis:

- **High-Risk Merchants**: Analysis reveals that fraud is not evenly distributed across merchants. Certain merchants consistently show fraud rates 2–3 times higher than average. These merchants represent key targets for fraud prevention efforts.

- **Merchant Risk Score**: An engineered feature, merchant risk score, has proven to be highly effective in detecting fraud patterns associated with specific merchants. This score captures historical fraud data and enhances model predictions.

2. Time-Based Fraud Patterns:

- **Fraud Peaks at Specific Times**: Fraudulent transactions tend to spike between 10:00 PM and 11:00 PM, as well as during weekends (Friday to Sunday**)**. This provides actionable information for designing fraud prevention systems that operate more efficiently during these peak times.

- **Seasonal Patterns**: Fraud detection analysis shows that fraud occurrences are higher in colder months, especially in January and February. Seasonal trends in fraud should be factored into predictive models and monitoring systems for more targeted fraud prevention.

3. Transaction Amount and Category Insights:

- **Transaction Amount**: Fraudulent transactions generally involve smaller transaction amounts compared to legitimate transactions, which often have larger outliers. The mean transaction amount is significantly different between fraud and non-fraud cases, helping to fine-tune fraud detection models.

- **Category Vulnerability**: Categories like Shopping, Misc_net, and Grocery_pos experience higher fraud rates. These categories should be monitored more closely to identify suspicious activity quickly.

4. Feature Engineering and Model Performance:

- **Feature Importance**: The most critical predictors of fraud include transaction amount, merchant risk score, time of transaction, and location differences (geographic discrepancies between customer and merchant).

- **Model Performance**: The Random Forest model performed exceptionally well, with an F1-score of 0.47 and an ROC-AUC score of 0.993, achieving 91% recall for fraud detection. This model effectively captures fraudulent transactions with minimal false negatives.

Strategic Recommendations:

1. **Focus on High-Risk Merchants**:

- **Targeted Fraud Prevention**: Given that fraud is concentrated around certain merchants, we recommend implementing merchant-specific fraud detection strategies, including monitoring merchants with high fraud rates (e.g., over 2%).

- **Merchant Risk Scoring**: Integrate the merchant risk score into existing fraud detection systems to provide a historical fraud risk for each transaction, improving model accuracy.

2. **Optimize Detection Based on Time**:

- **Time-Sensitive Alerts**: Enhance fraud detection systems by scheduling real-time alerts for transactions that occur between 10 PM and 11 PM and over weekends. Fraudulent activity tends to peak during these times.

- **Seasonal Monitoring**: Implement seasonal fraud detection strategies during colder months, especially January and February, to reduce the chances of missed fraudulent transactions.

3. **Transaction-Level Monitoring**:

- **Amount-Based Alerts**: Transactions of smaller amounts should be flagged as potential fraud, especially if they fall within specific categories (e.g., Shopping and Grocery_pos).

- **Category-Based Monitoring**: Focus fraud detection efforts on categories that are more prone to fraud, such as Shopping and Misc_net, to ensure resources are allocated where needed most.

4. **Model Improvement and Refinement**:

- **Hyperparameter Tuning**: Future efforts should focus on tuning the Random Forest model or experimenting with other ensemble techniques to further reduce false positives while maintaining high fraud detection recall.

- **Consider Hybrid Models**: To enhance fraud detection capabilities, consider using hybrid models that combine Random Forest, Logistic Regression, and even Deep Learning approaches, capturing more complex patterns in the data.

Conclusion:

The analysis highlights that credit card fraud is not randomly distributed, but rather concentrated in specific merchants, times, and transaction amounts. By leveraging these insights, businesses can optimize fraud detection models, improve real-time monitoring, and enhance fraud prevention systems. The merchant risk score is a particularly powerful feature that can significantly improve fraud detection accuracy.