From Victim to Defender: Using Predictive Analytics to Detect Financial Fraud

- Hasnaa Elidrissi
- Department of Data Science, Bellevue University

- DSC 630: Predictive Analytics

Introduction

Personal story of financial fraud experience.

The growing problem of fraud and its financial impact.

Importance of predictive analytics in detecting fraudulent transactions.

Problem Statement

Financial fraud is a global issue, costing businesses trillions annually.

Manual fraud detection is ineffective due to transaction volume.

Need for machine learning techniques to improve fraud detection.

Dataset Overview

- Kaggle dataset containing labeled financial transactions.
- Imbalance issue, only 0.13% of transactions are fraudulent.
- Feature include transaction amount, type, balance changes.

Data Preparation

One-hot encoding for categorical variables.

Class imbalance managed through class weighting.

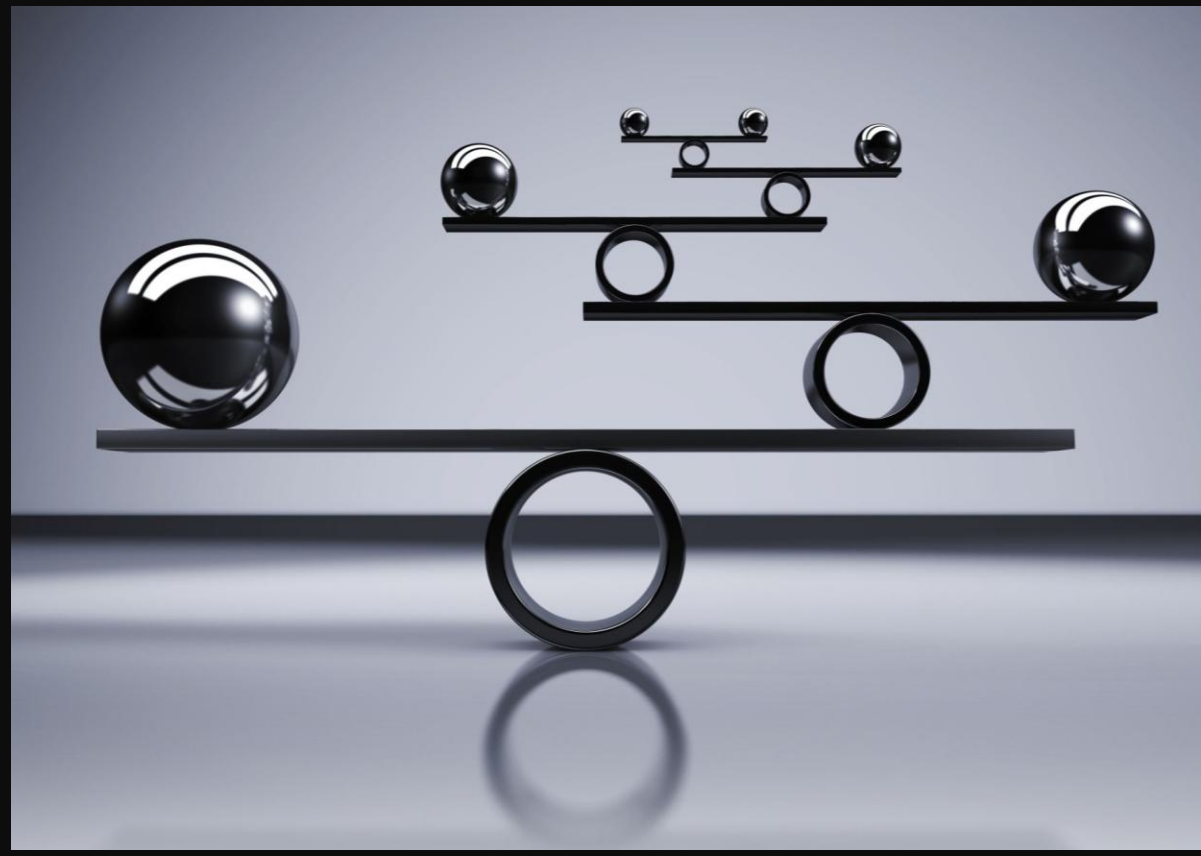
Feature selection to remove unnecessary data.

Machine learning Models

Random Forest	XGBoost	Gradient Boosting & Logistic Regression
Best balance of precision and recall.	High recall by too many false positives.	Lower recall rates.

Model Evaluation Metrics

- Precision: Accuracy in identifying fraudulent transaction.
- Recall: Ensuring fraud cases are detected.
- F1-Score: Balancing precision and recall.
- ROC-AUC: Overall performance across classification thresholds.



Results & Model Comparison

Random Forest: 96% precision, 78% recall.

XGBoost: Detected almost all fraud but too many false positives.

Gradient Boosting & Logistic Regression: Missed too many fraud cases.

Key Insights

- Fraud is concentrated in ‘Payment’ and ‘cash_out’ transactions.
- Smaller transaction amounts correlate with fraudulent activity.
- Random Forest provides the best trade-off between detection and false positives.
- Importance of balancing fraud detection with financial impact and user experience.

Conclusion & Model Recommendation

- Random Forest chosen as the best model.
- XGBoost is promising but needs further tuning.
- Fraud detection requires balancing precision and recall effectively.

Final Takeaways

Balancing precision and recall is crucial for effective fraud detection.

Fine-tuning models can significantly improve accuracy and reduce false positives.

Larger datasets and continuous learning enhance fraud detection systems.

Financial institutions must minimize fraud while reducing friction for legitimate users.

References

- Association of Certified Fraud Examiners (ACFE). (2022). Report to the Nations: Global Study on Occupational Fraud and Abuse. Retrieved from https://www.acfe.com
- Aman. (n.d.). Financial Dataset for Fraud Detection in a Company. Retrievedfromhttps://www.kaggle.com/datasets/amanindiamuz/financial-dataset-for-fraud-detection-in-a-company