PREDICTIVE FRAUD ANALYTICS

IMPROVING EARLY FRAUD DETECTION WHILE PROTECTING CUSTOMER EXPERIENCE

PRESENTED BY: MARGARET MONDIA

DECEMBER 2025

INTRODUCTION

This project builds a predictive model to identify high-risk transactions before loss occurs.

Why this matters:

- Credit card fraud drives significant financial losses
- Overly strict controls disrupt legitimate customers
- The challenge is detecting fraud early without harming trust

BUSINESS PROBLEM

The main problem surrounding this business are:

- Fraud losses continue to rise across digital payments
- Manual review capacity is limited
- Excessive false alerts increase cost and customer friction

The business therefore needs smarter prioritization, not blanket blocking.

PROJECT OBJECTIVE

This project:

- Identify the highest-risk transactions in real time
- Help fraud teams focus attention where it matters most
- Support decisions — not replace human judgment

This analysis addresses when fraud can be detected early, when human review is required, how much fraud can be captured, and how transaction value informs risk prioritization.A

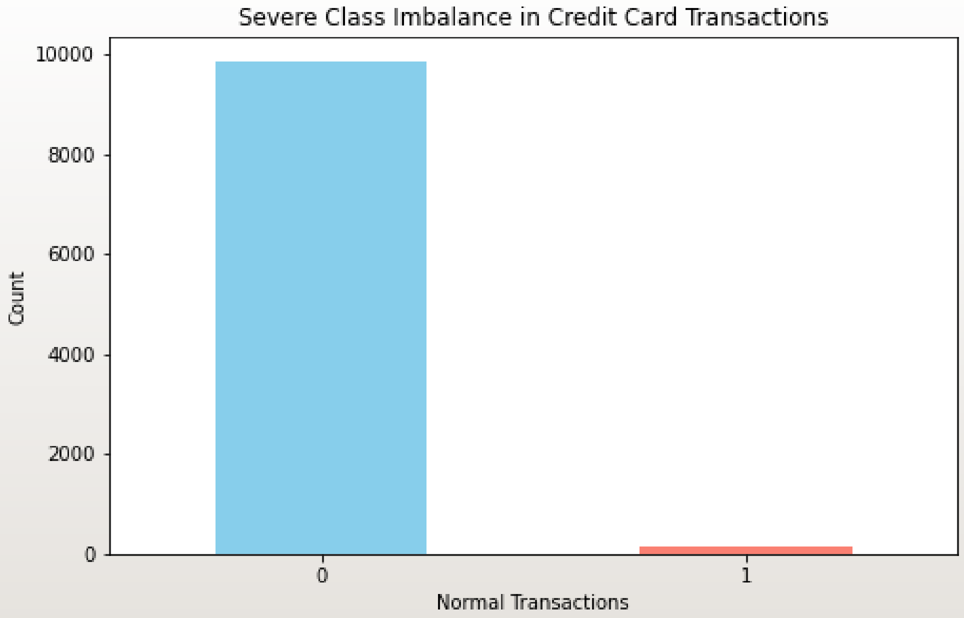
BUSINESS UNDERSTANDING

1. Primary stakeholder: Fraud Prevention Team
2. Cost of missed fraud far exceeds false alerts
3. Recall is prioritized over overall accuracy

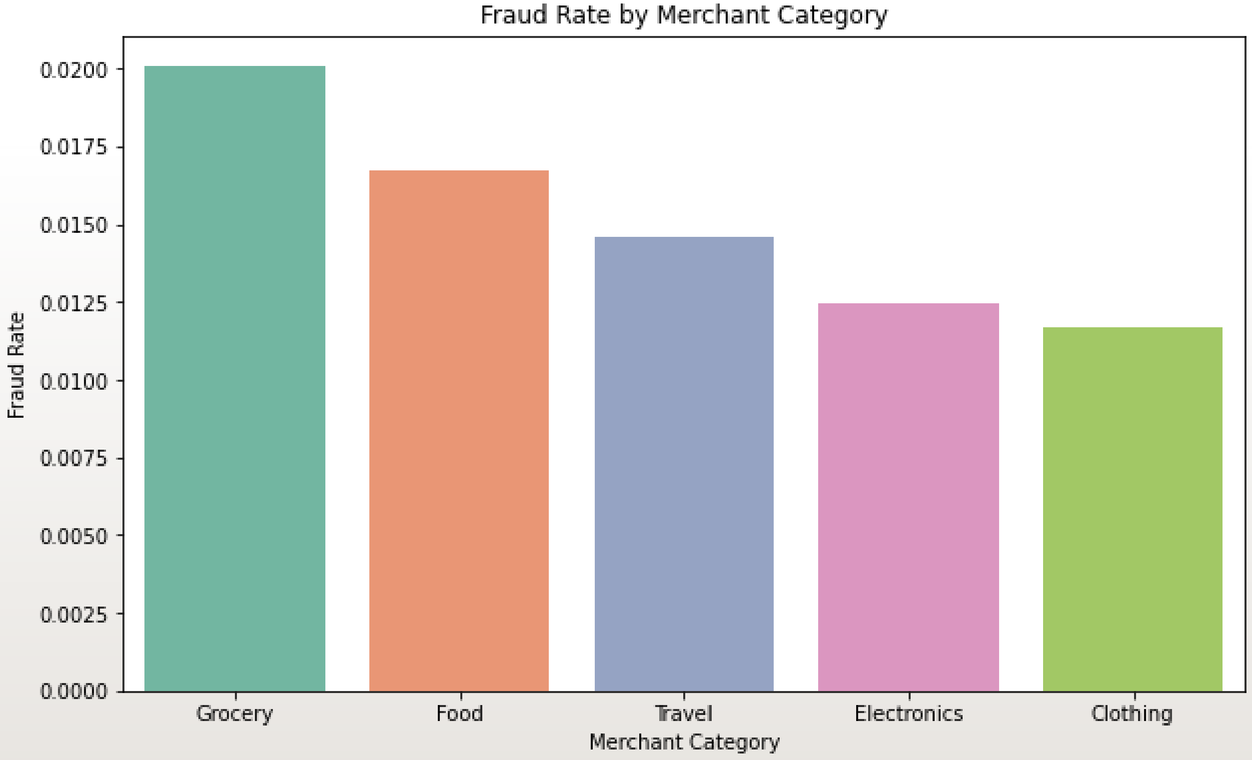
DATA OVERVIEW

Dataset Summary

- 10,000 real credit card transactions
- Fraud rate: ~1.5% (high class imbalance)
- Features include:
  - Transaction amount & timing
  - Device trust and transaction velocity
  - Location and merchant category
- This closely reflects real-world fraud conditions.

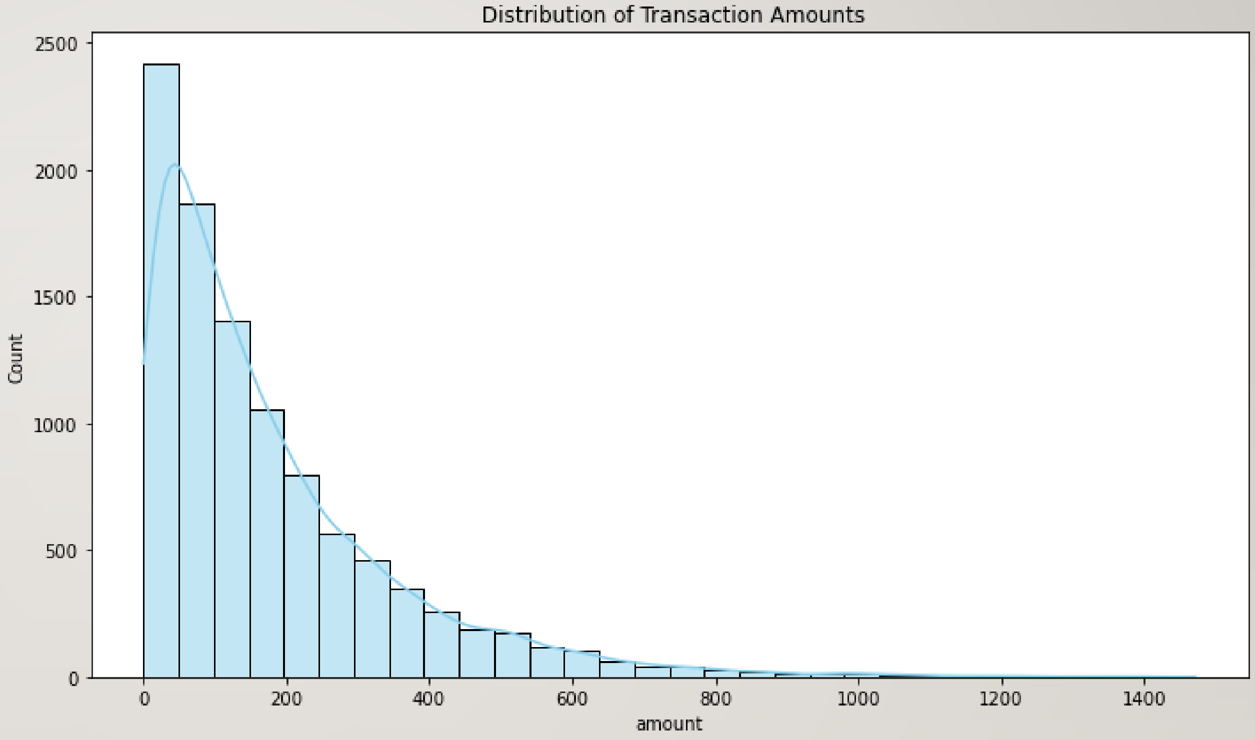


KEY DATA INSIGHTS



Certain merchant categories show elevated fraud rates, indicating opportunities for category-specific risk controls and targeted monitoring.

KEY DATA INSIGHTS



Transaction amounts are heavily right-skewed, with most transactions occurring at lower values and a long tail of high-value transactions

MODELING APPROACH

Model strategy

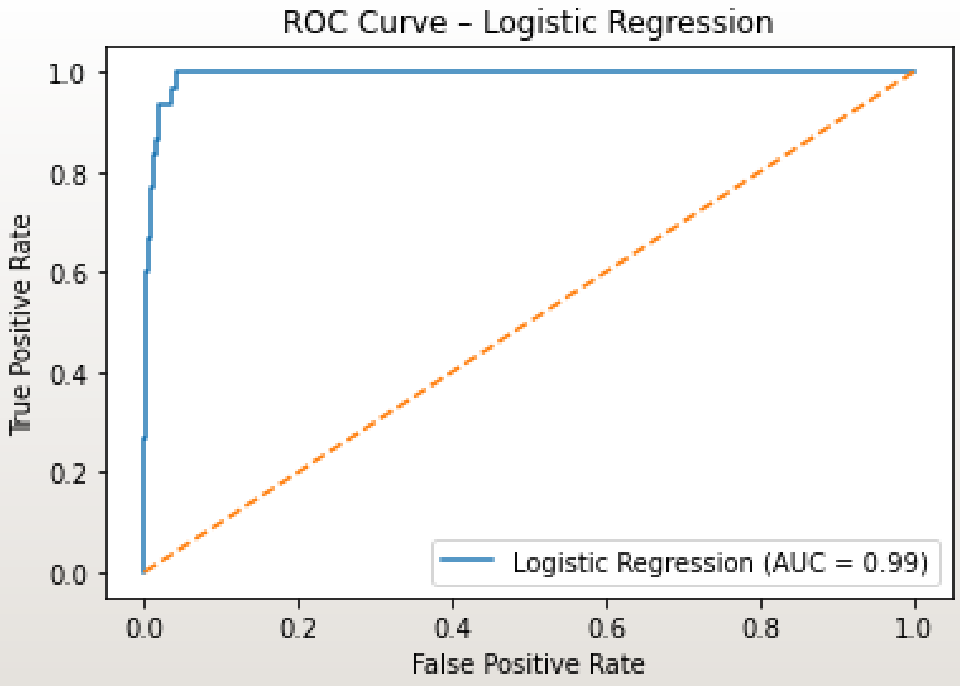
- Logistic Regression for baseline interpretability
- Decision Tree to capture non-linear risk patterns
- Class imbalance handled through weighted learning

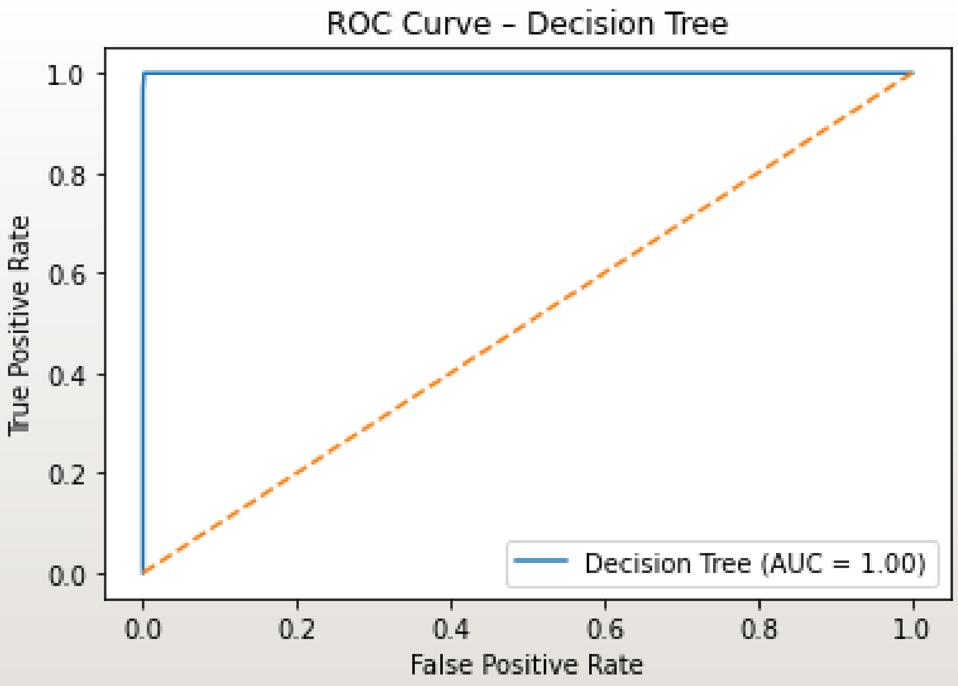
How Performance Was Measured

- Recall prioritized to minimize missed fraud
- ROC-AUC used to assess risk separation
- Precision monitored to manage customer impact

Evaluation focused on business relevance, not technical vanity metrics.

MODEL RESULTS





- Both models detect nearly all fraud cases
- Decision Tree improves precision and risk separation

MODEL SUMMARY

Performance Summary

- Both models detected nearly all fraudulent transactions
- Decision Tree:
  - Higher precision than Logistic Regression
  - Better separation of risky vs safe transactions
  - Stronger overall risk ranking

Decision Tree selected as the final model.

KEY FINDINGS

What Drives Fraud Risk:

- Foreign or unusual locations → ~40× higher risk
- High transaction frequency → strong fraud indicator
- Low device trust → potential stolen card usage
- Electronics and Clothing merchants show higher fraud exposure

- Foreign or unusual locations greatly increase risk
- High transaction frequency indicates fraud testing

BUSINESS VALUE AND REVIEW

This enables smarter, risk-based decision making by the business as it:

- Enables real-time transaction screening
- Flags top 1–2% highest-risk transactions
- Improves analyst efficiency and customer experience

The model enhances decisions and does not replace people.

Human Judgment Remains Essential include:

- Borderline probability cases
- High-value transactions
- Customer-sensitive scenarios

RECOMMENDATIONS

1. Integrate model into real-time fraud monitoring
2. Adjust risk thresholds dynamically during high-risk periods
3. Apply stricter controls to high-risk merchants and transactions
4. Monitor performance continuously

LIMITATIONS

1. Performance reflects dataset simplicity
2. Real-world fraud is more complex and constantly evolving
3. Continuous monitoring and retraining required

FUTURE ENHANCEMENTS

Opportunities for Improvement include

- Ensemble models (Random Forest, Gradient Boosting)
- Threshold optimization based on business capacity
- Regular retraining to adapt to new fraud patterns

CONCLUSION

1. The Decision Tree model effectively separates risky from safe transactions
2. It improves early fraud detection while protecting customer experience
3. Provides a strong foundation for scalable, risk-based fraud prevention

THANK YOU!

Margaret Mondia

Data Scientist

margaret.mondia@student.moringaschool@gmail.com

https://github.com/megmondia-spec