DSCI478 Kaggle Project - Credit Card Fraud Detection

Nick Brady, Jakob Wickham

February 24, 2025

Note for us: - If you want the PDF to not display a cell, click the three dots on the cell, click "Add Cell Tag", and put "remove_cell" - If you want to hide the code instead, put "remove_input"

1 Introduction

Talk about the importance of fraud detection and fraud prevention

State the dataset was synthetically generated

2 Data Cleaning and Preprocessing

Dropping the TX_FRAUD_SCENARIO column and turning the TX_DATETIME column to a proper DateTime

3 Exploratory Data Analysis (EDA)

General stuff, like fraud percentage, rolling window sample distributions, etc

Graphs such as ratio of fraud within sliding windows, and then some

4 Feature Engineering

Imbalanced techniques, adding columns for rolling windows and sampling distribution z-scores, etc

5 Model Selection

Isolation Forest, Balanced Random Forest, etc

6 Model Training and Evaluation

7 Model Interpretation and Explainability

Graphs such as feature importance and within-feature / within-fraud boxplots

8 Conclusion

Talk about the three scenarios that we used to make a transaction fraud

9 References

- Reproducible Machine Learning for Credit Card Fraud Detection - Practical Handbook
  - Yann-Aël Le Borgne, Wissam Siblini, Bertrand Lebichot, and Gianluca Bontempi
  - https://github.com/Fraud-Detection-Handbook/fraud-detection-handbook