A Fraud Screener for American Express

Chuck Nadel

Overview

- Problem: American Express is looking to be more effective at screening for fraud

- Task: Create a classification algorithm to screen for fraudulent purchases

- Goal: Maximize rate of fraudulent purchases caught

Agenda

1. Data Overview

2. Data Exploration

3. Model Overview

4. Results

5. Next Steps

Data Overview

Credit Card Transactions Fraud Detection Dataset

- Dataset via Kartek Shenoy on Kaggle

- Simulation of Credit Card Transactions between 1000 people and 800 potential vendors
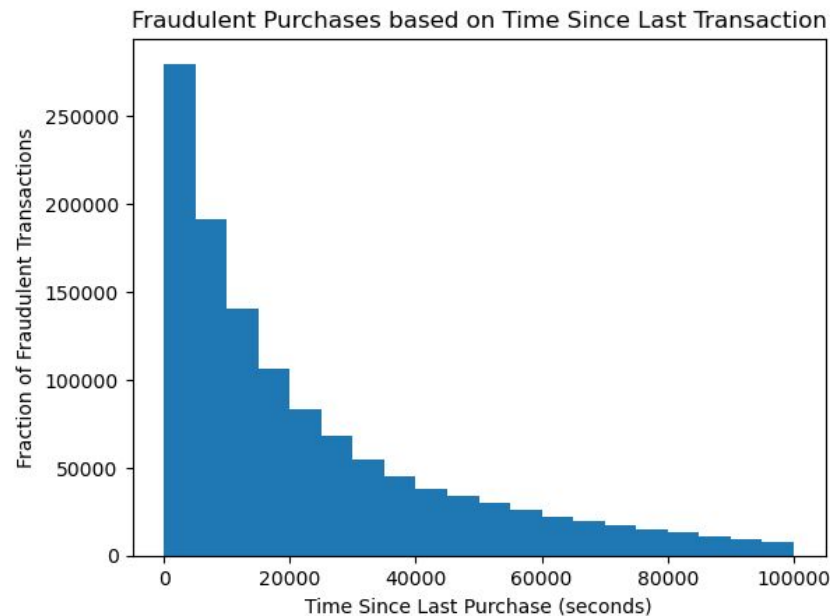
Simulation Details

- Ran using Brandon Harris' Sparkov Data Generation Github from January 1st, 2019 to December 30th, 2020

- Tracked 22 attributes for 1.85 million transactions, as well as whether or not each purchase was fraudulent

Data Exploration

Distance from Home

Histogram of Fraudulent Purchases by Distance

Time Since Last Purchase

Fraudulent Purchases based on Time Since Last Transaction
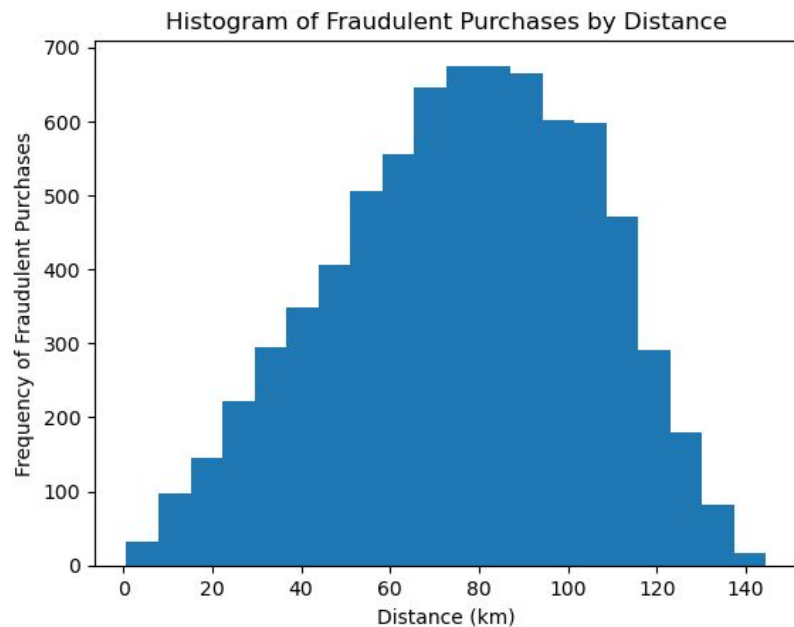
Model Overview

First Model: Logistic Regression
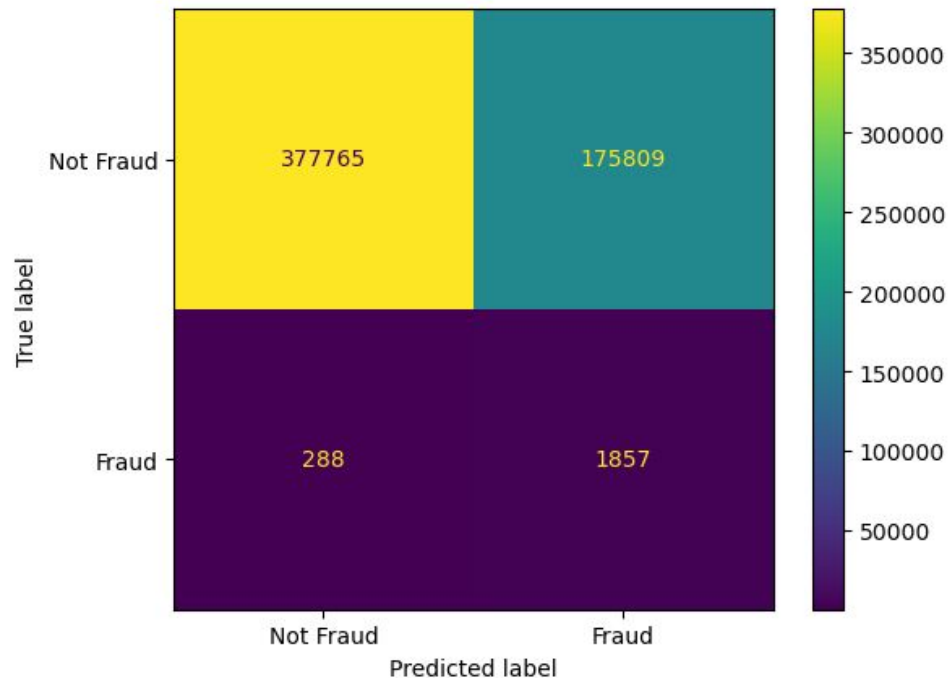
Second Model: Decision Tree

Predictive Variables:

- Purchase Amount
- Gender
- City Population
- Category of Purchase
- Distance from Home
- Time Since Last Purchase

Logistic Regression

Optimized for Custom Recall Score (Threshold = 0.3)

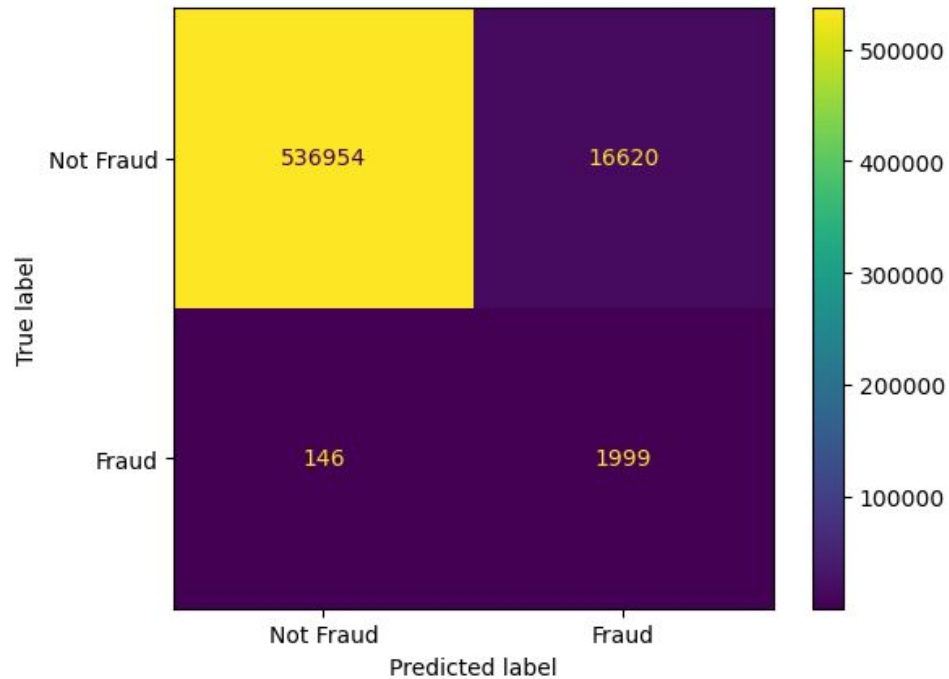Recall = 0.87
False Positive Rate = 0.31

Decision Tree Model

Optimized for Recall Score
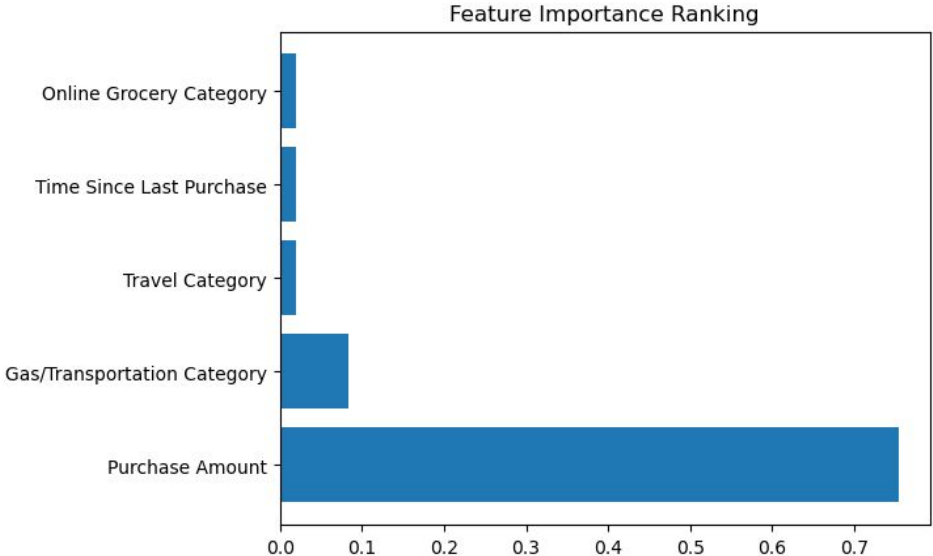
Recall = 0.93
False Positive Rate = 0.03

Caught 98.6%

of Money Fraudulently Spent

Feature Importance Ranking

Feature Importance Rankings

Feature Name	GINI Importance
Purchase Amount	75.6%
Gas/Transportation Category	8.29%
Travel Category	2.00%
Time Since Last Purchase	1.98%
Online Grocery Category	1.92%

Next Steps

- Create New Scorer to Optimize on based on a mix of Recall, False Positive Rate

- Feature to Measure Distance Travelled between past 2 purchases

- Test on Real World Data

Thank You!

Email: cnadel14@gmail.com

Github Repository: https://github.com/ChuckNadel/ClassifyingCreditCardFraud

LinkedIn: https://www.linkedin.com/in/chuck-nadel-2ab034173/