

# Fraud Detection

# Problem

Fraud in online transactions:

- Loss of revenue
- Decreased customer satisfaction

# Solution

Fraud detection system:

- Accurate fraud prediction
- Less false alarms

# Data

- IEEE Computational Intelligence Society Fraud Detection
- Vesta Corporation's real-world e-commerce transactions
- 590,540 observations, 434 variables, 118 features

# Workflow

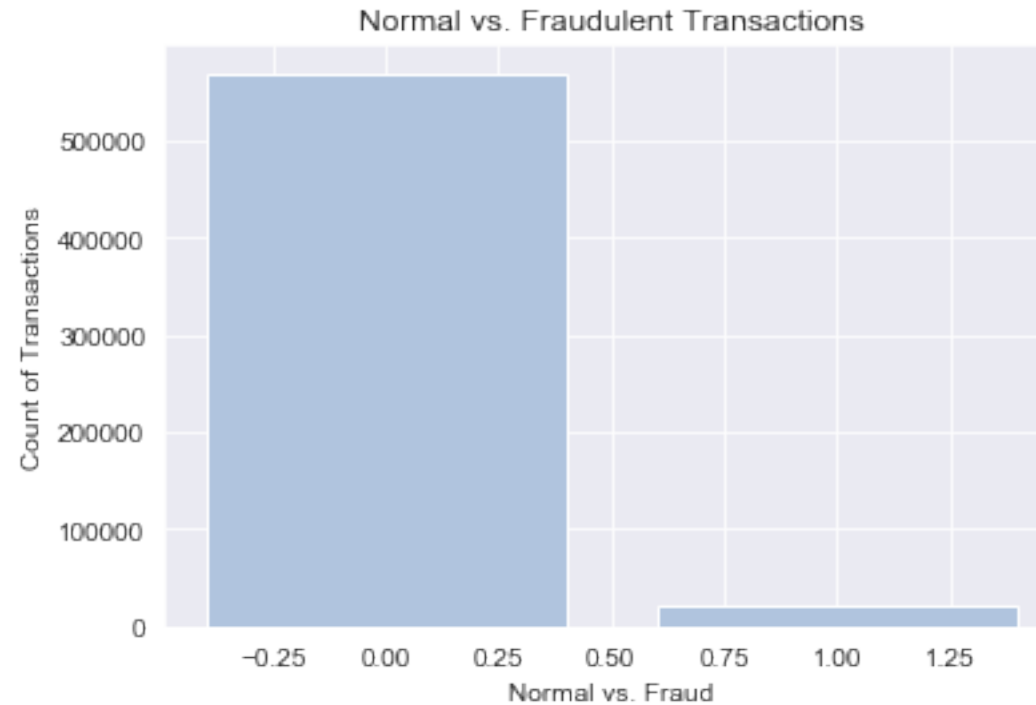
## I. Data Preparation:

- Data Cleaning
- Exploratory Data Analysis
- Feature Engineering

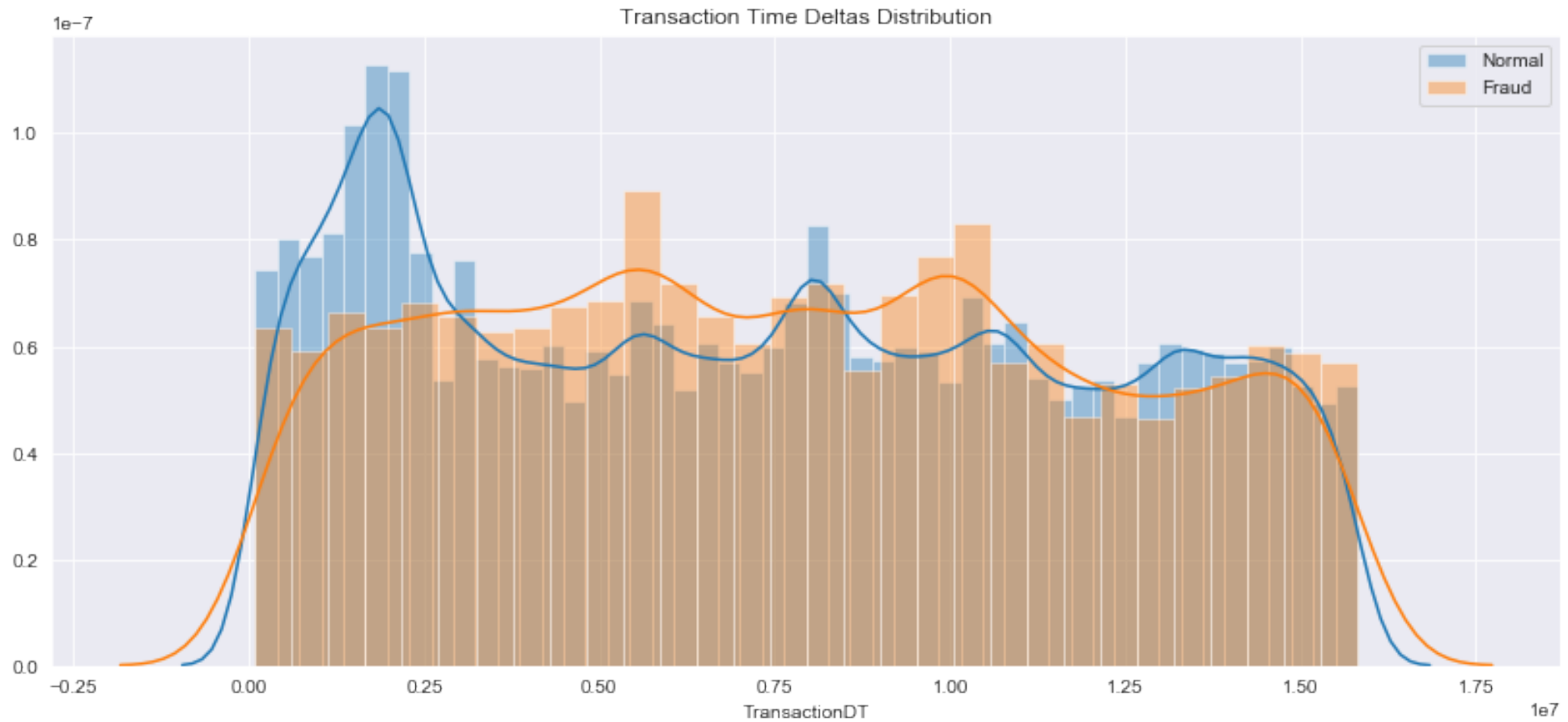
## II. Modeling:

- Class Imbalance
- Model Evaluation and Selection
- Hyperparameter Tuning

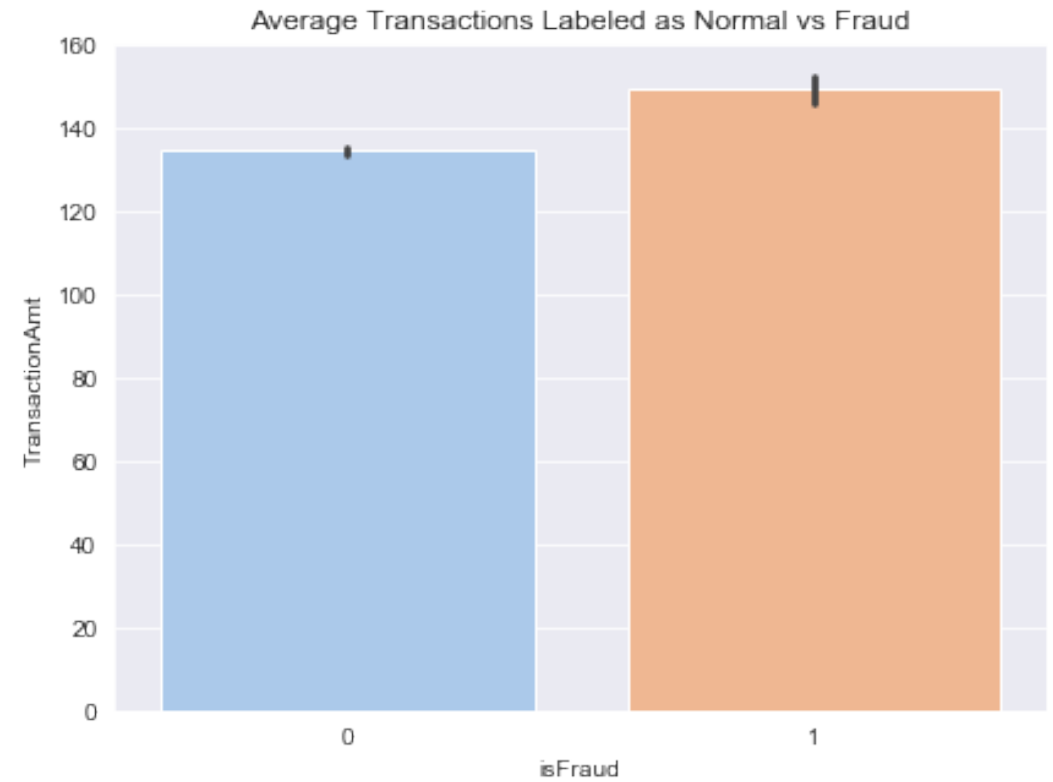
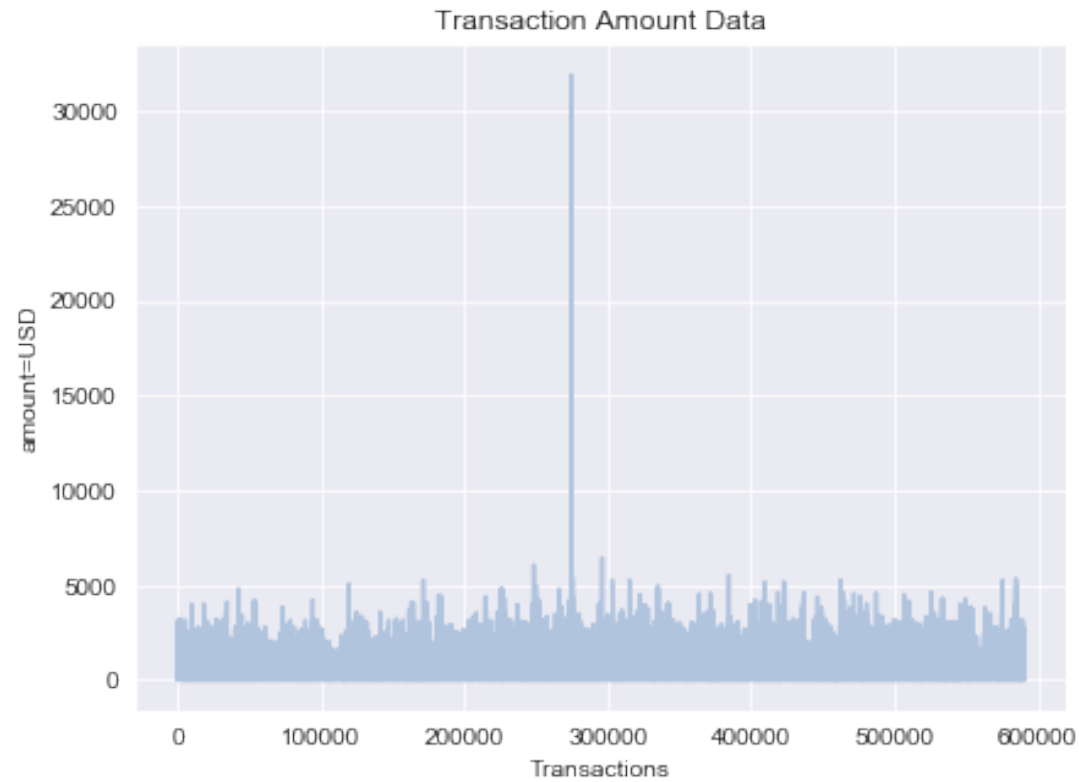
# Target Variable: class imbalance



# Transaction Time Deltas

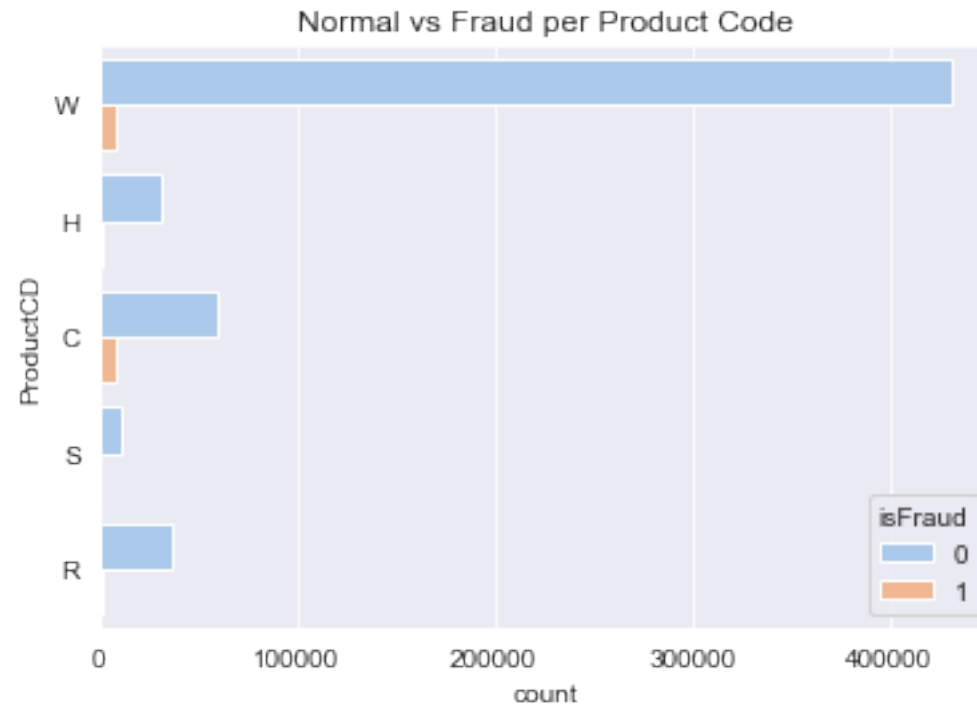


# Transaction Amount

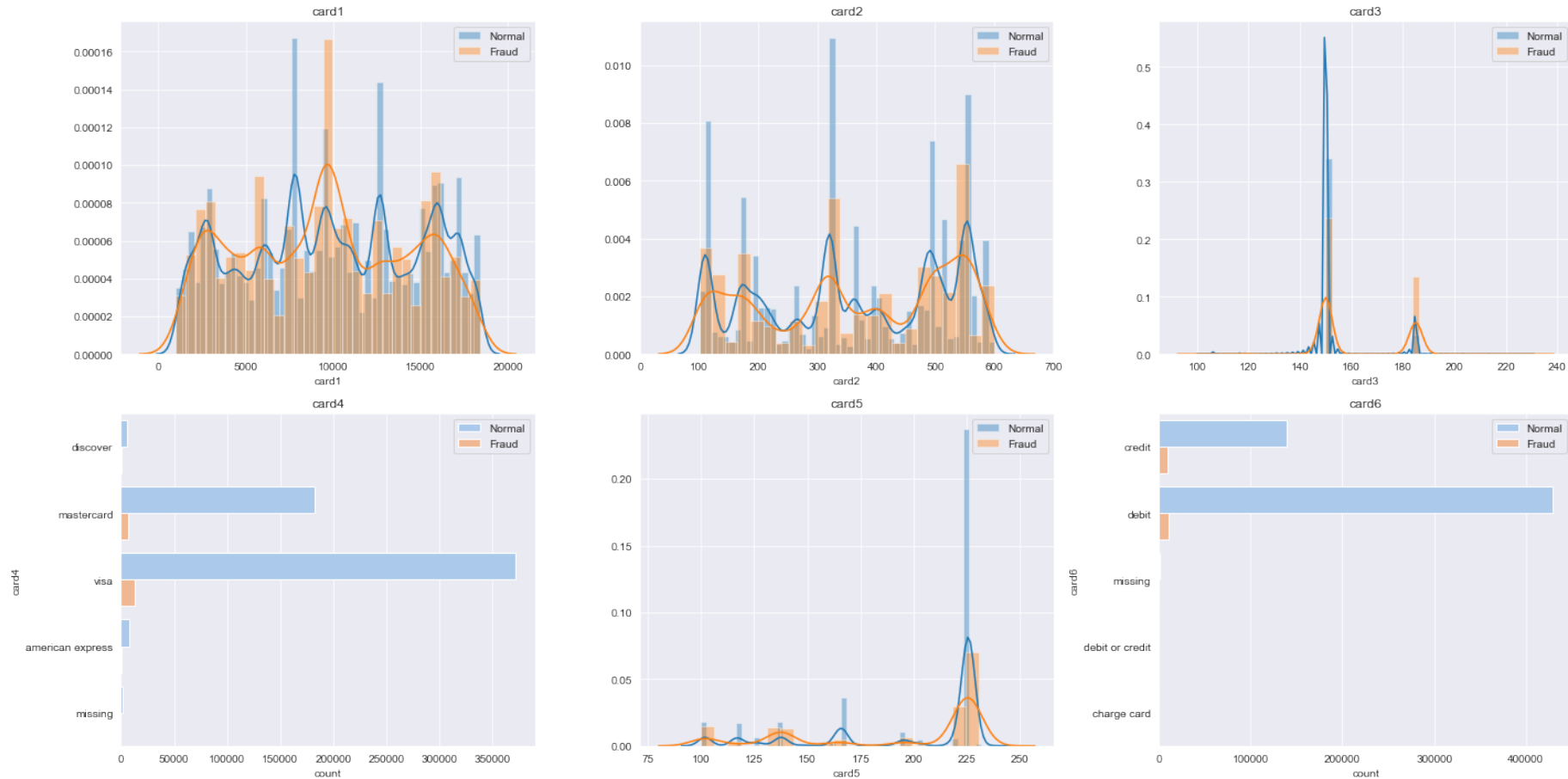




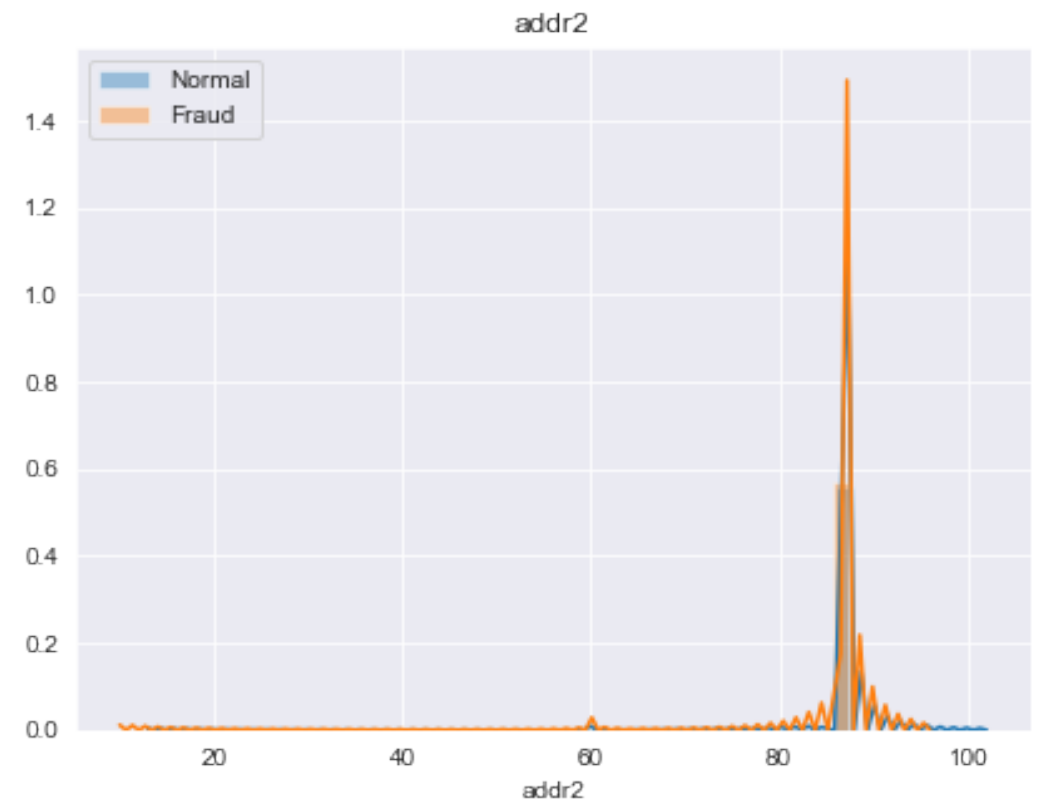
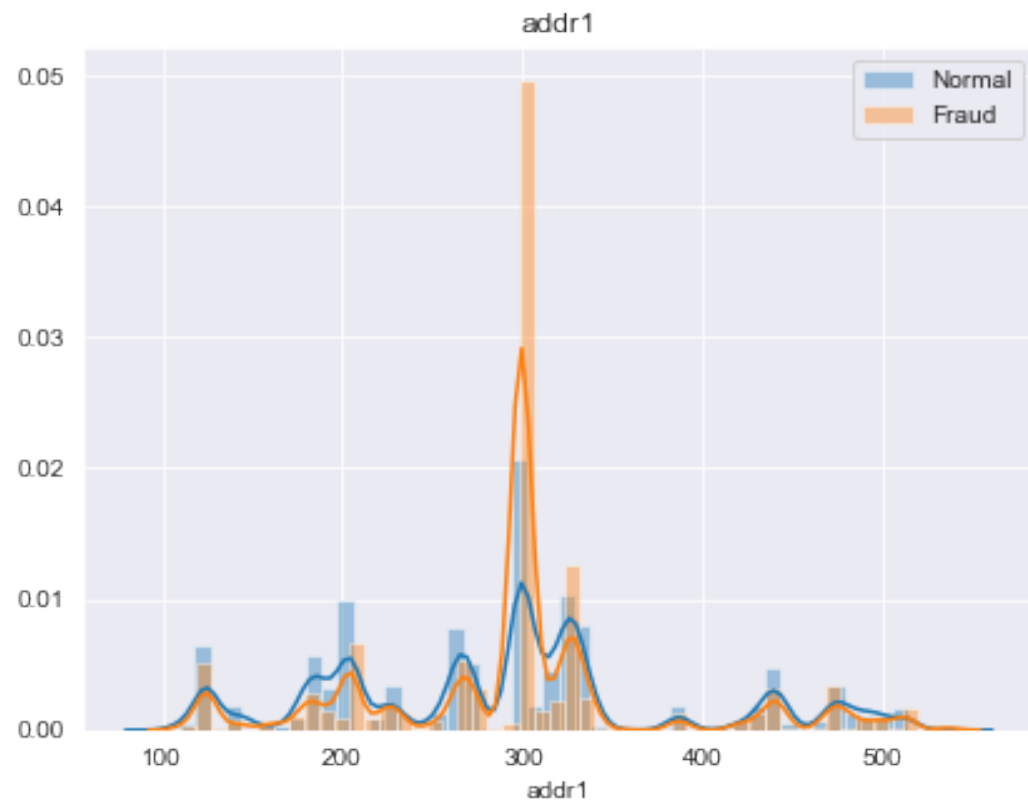
# Product Codes



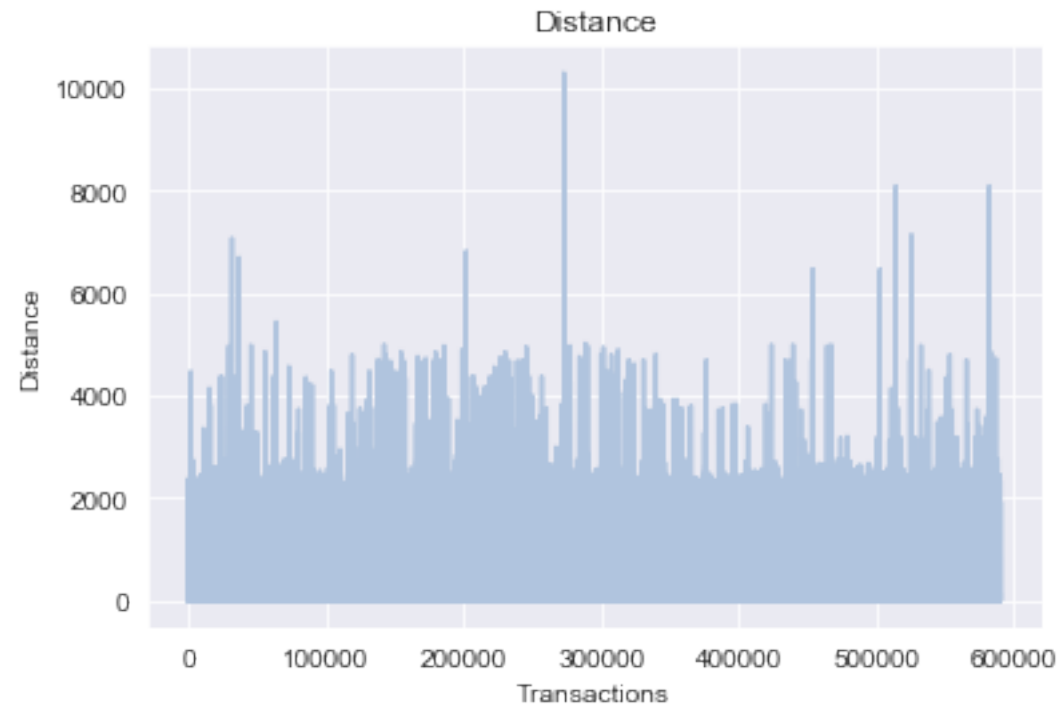
# Payment Card Information



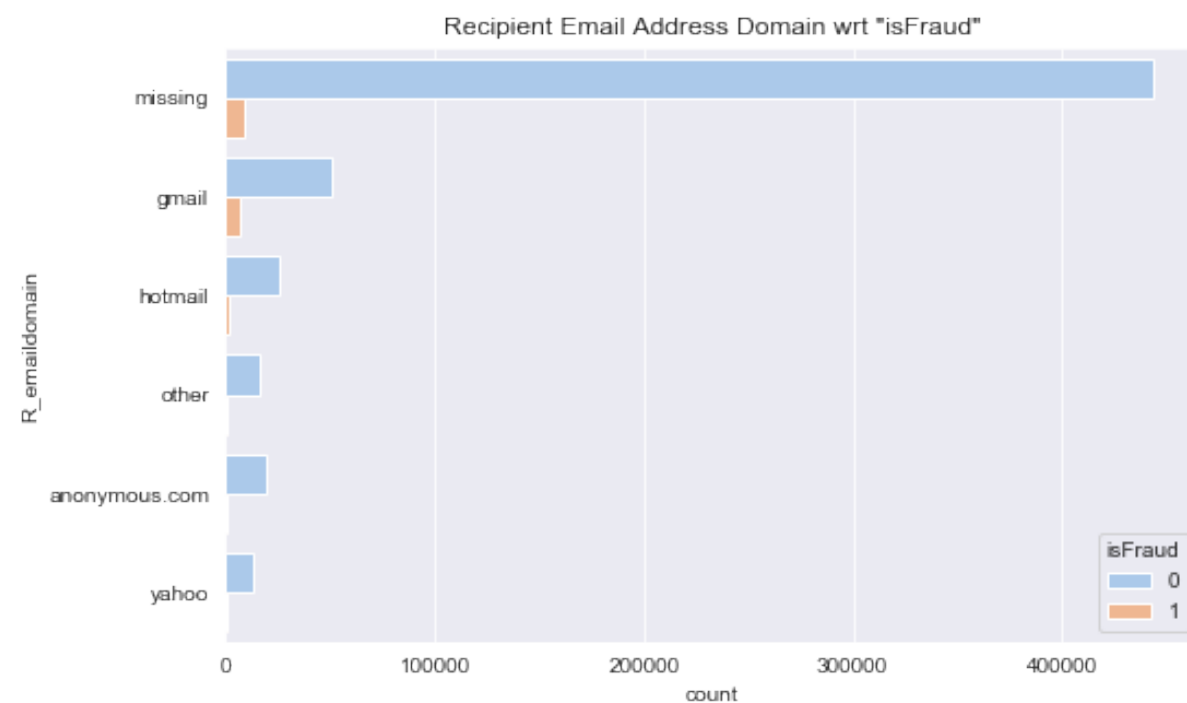
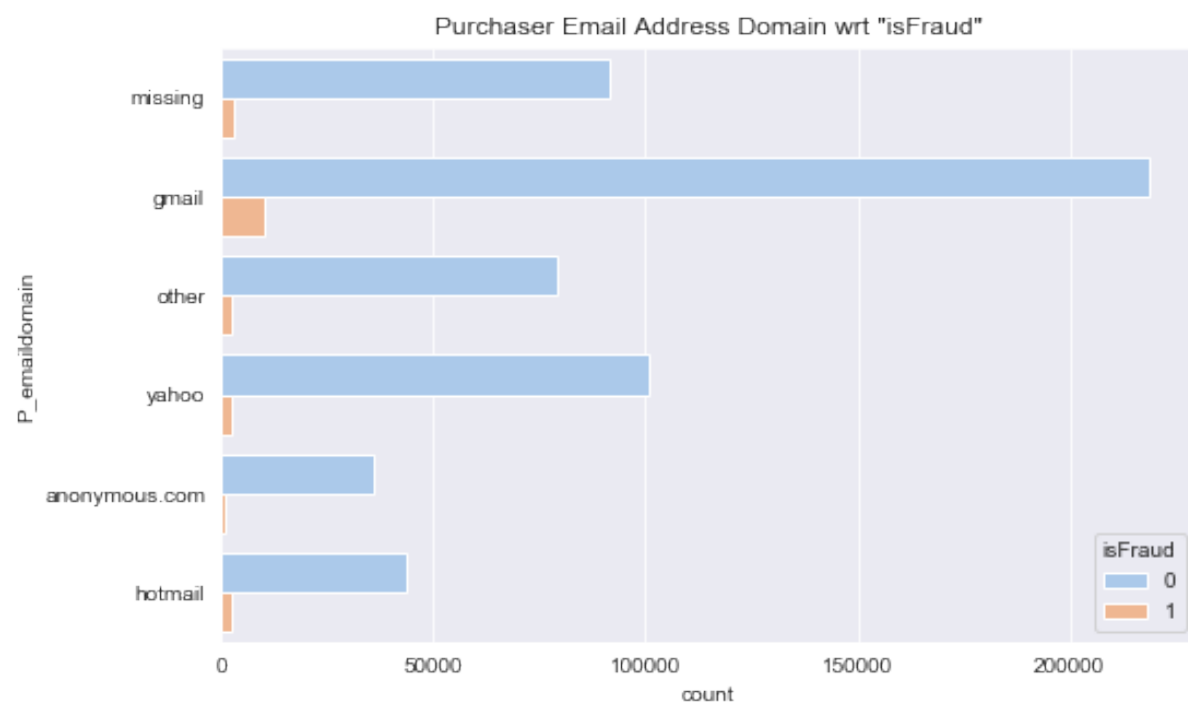
# Address



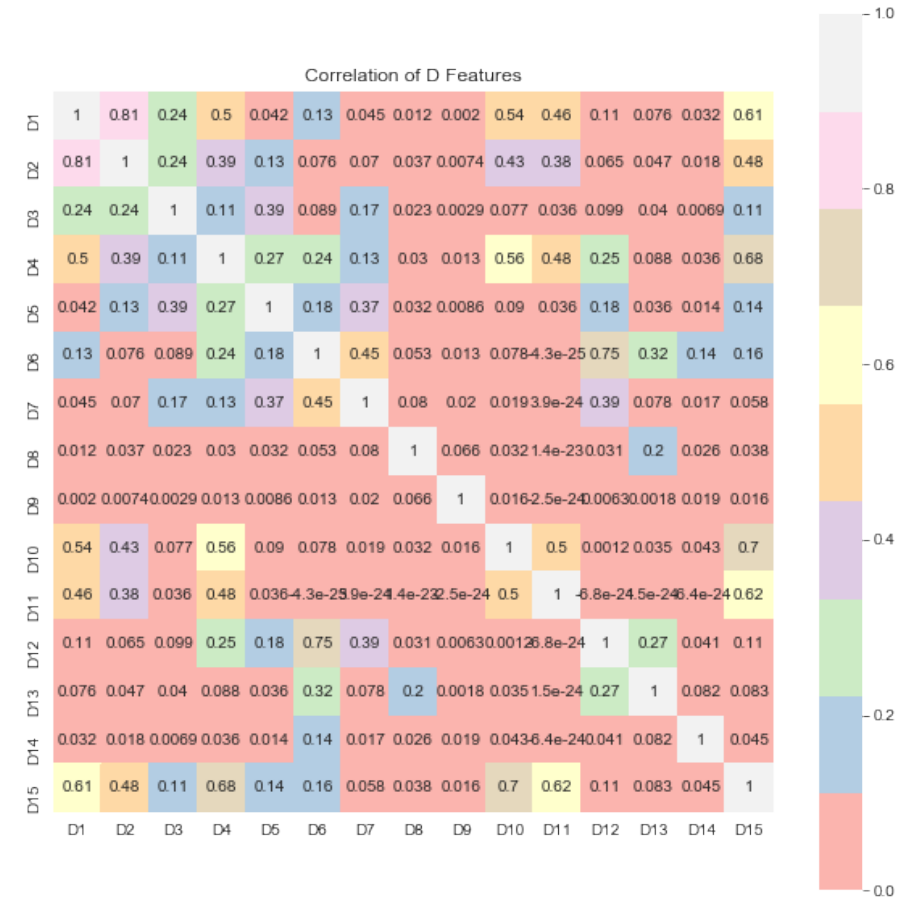
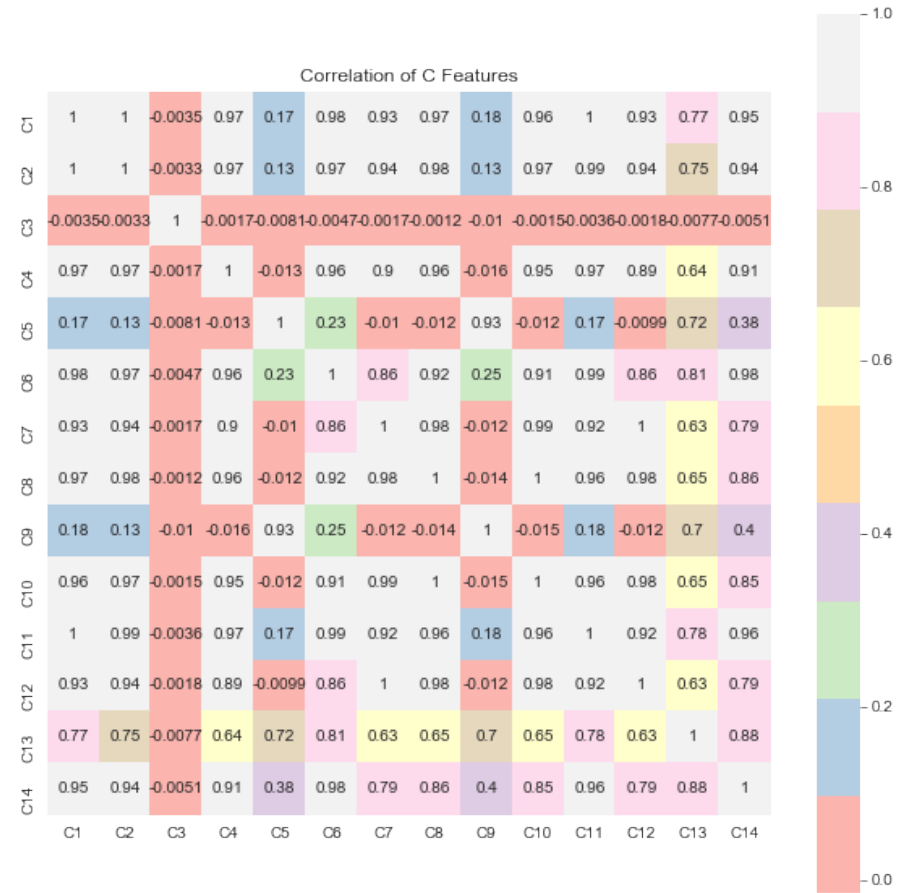
# Distance



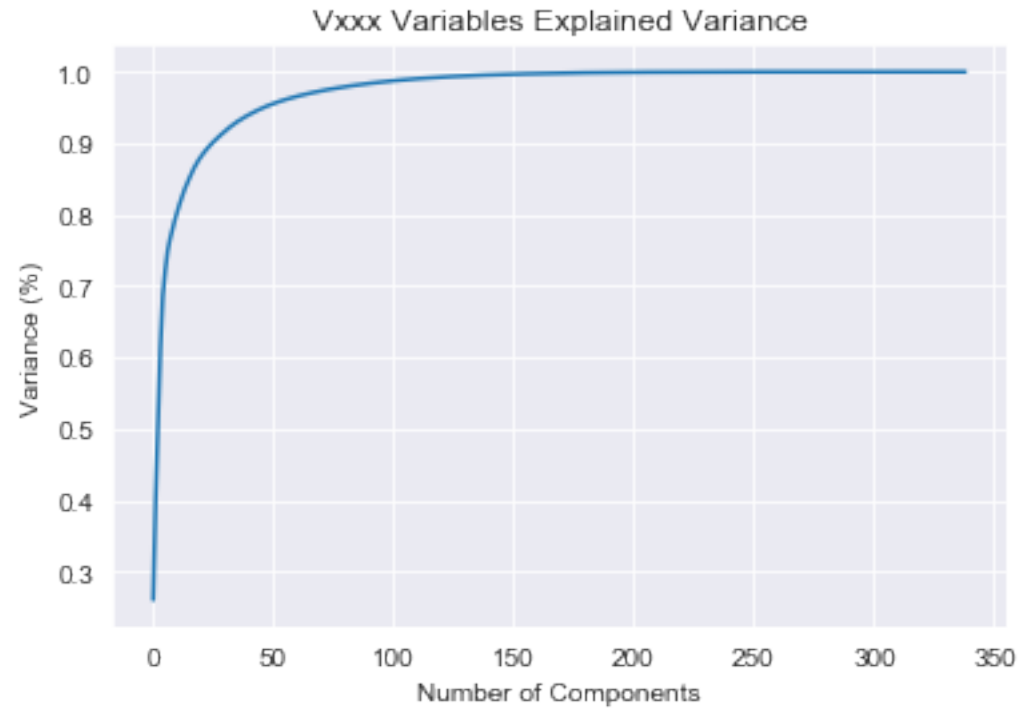
# Email Domains



# Feature Correlation

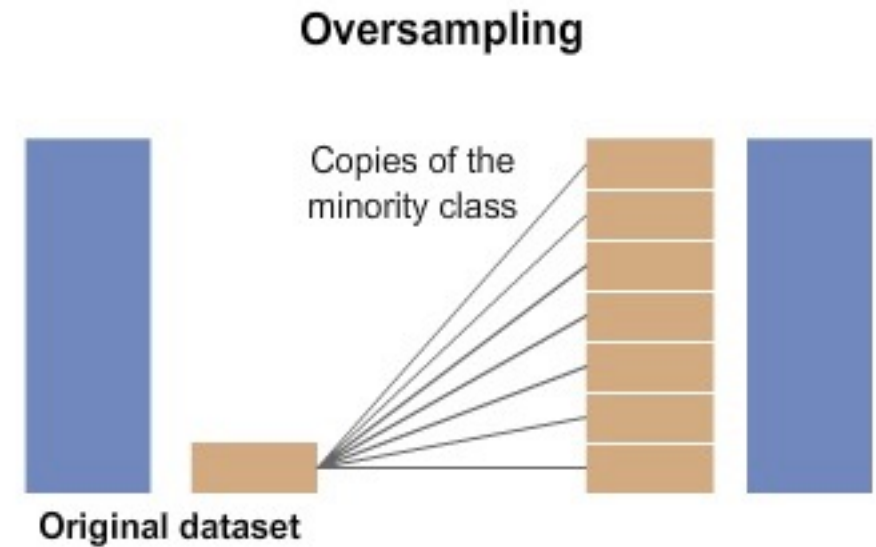
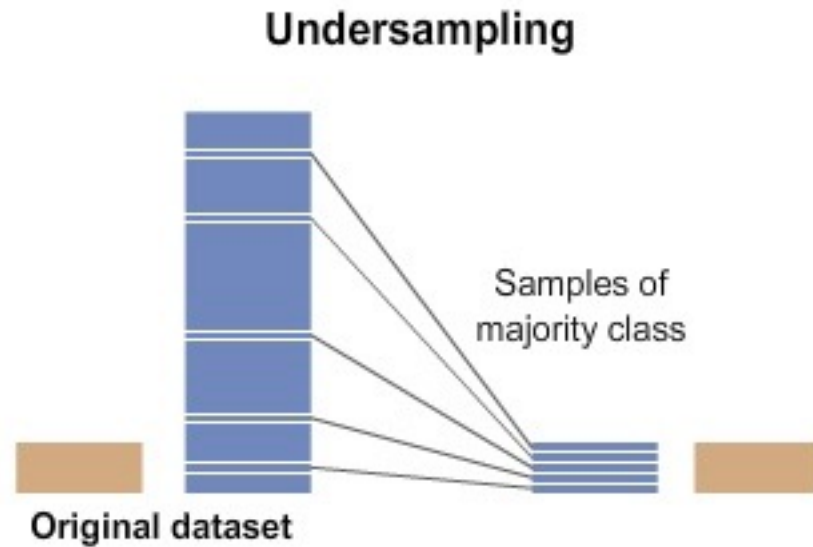


# Principal Component Analysis



# Class Imbalance

- Random Undersampling
- Synthetic Majority Oversampling Technique (SMOTE)

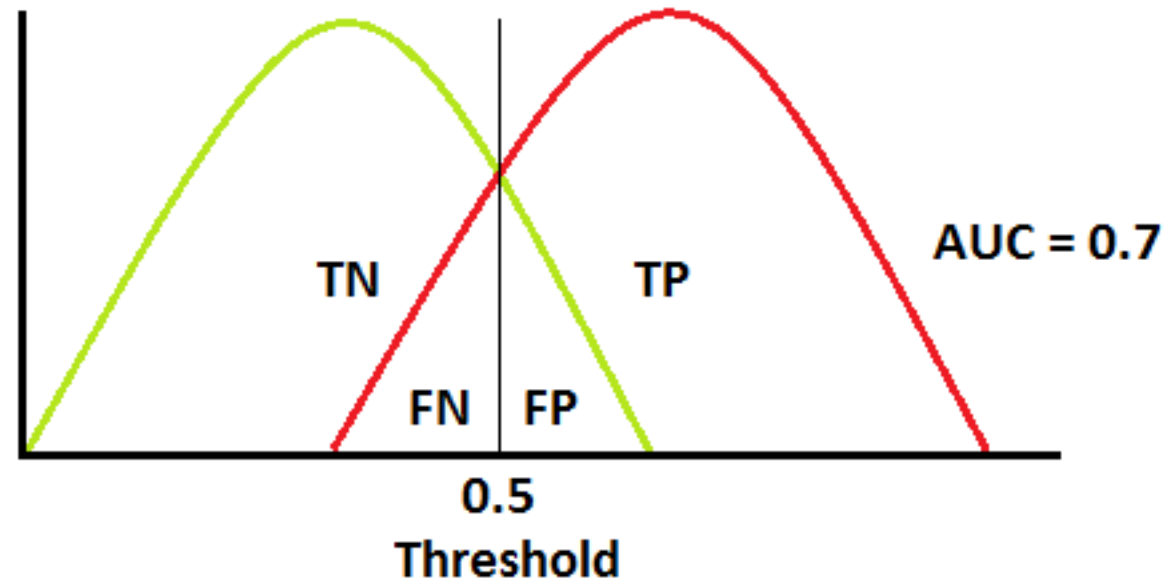


([source](#))



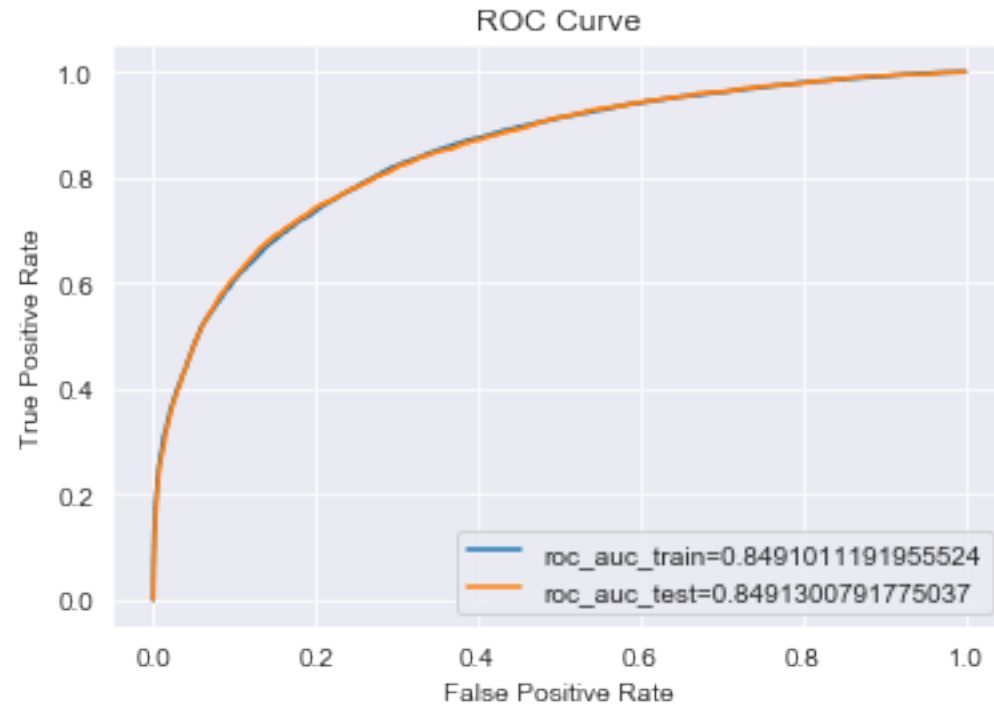
# Evaluation Metrics

- Area under Receiver Operating Characteristic curve
- 0.9

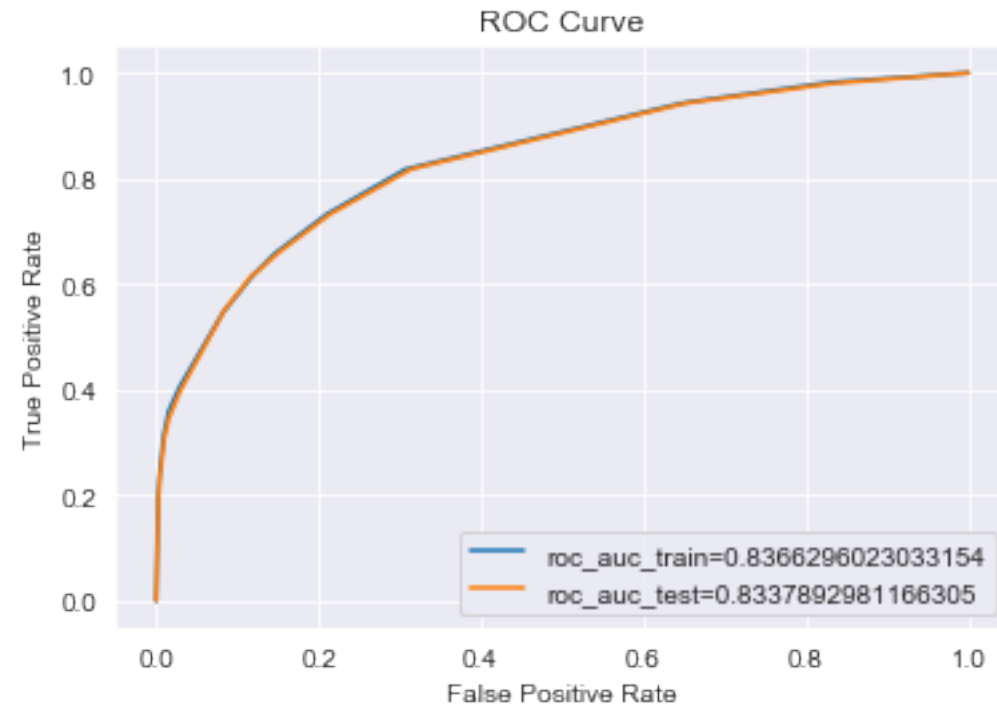


([source](#))

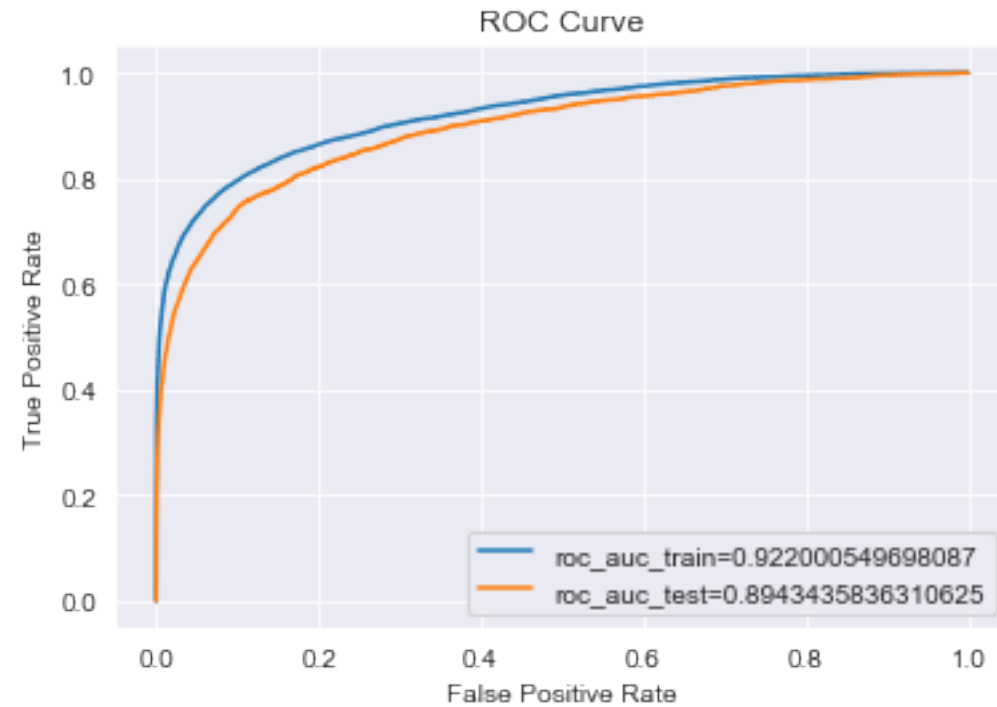
# Logistic Regression



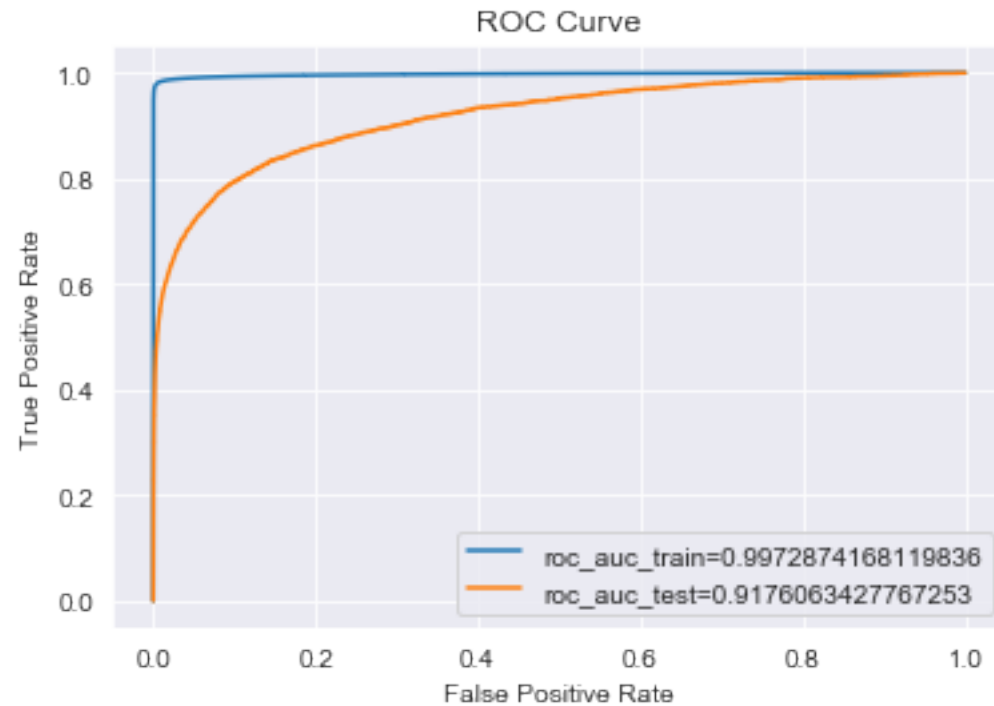
# Decision Tree



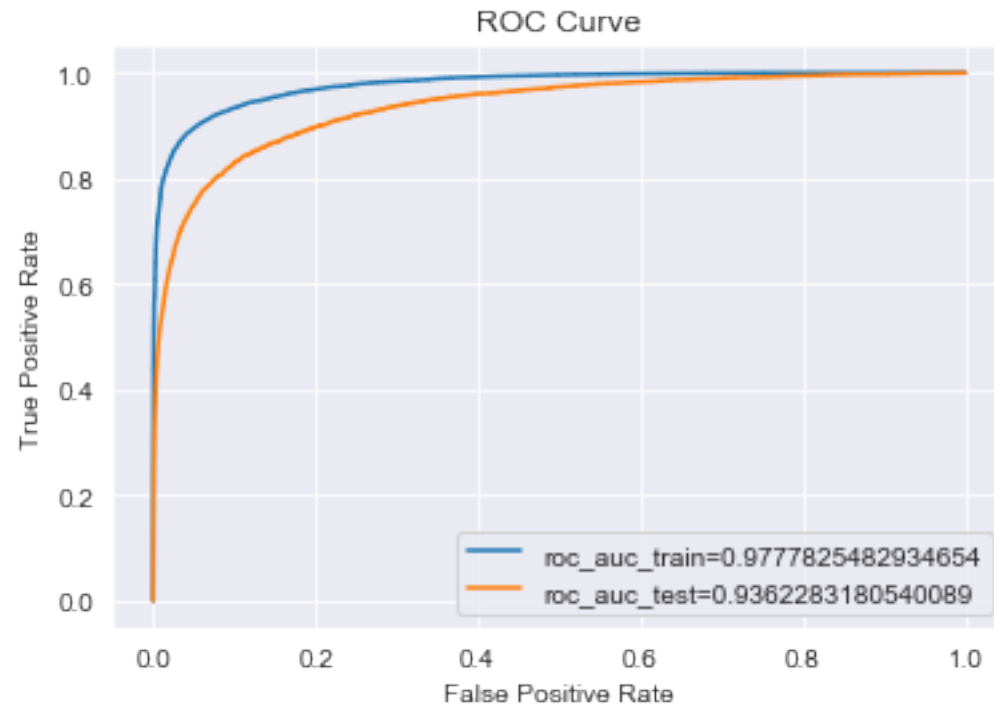
# Random Forest



# Gradient Boosting Classifier - SMOTE



# Gradient Boosting Classifier



# Important Variables ( $\geq 0.03$ )

- Vesta engineered features (actual meaning is masked):
  - I, PCA (e.g. ranking, counting, other entity relations)
  - C features (counting, e.g. how many addresses found)
  - D3 (time delta, such as days between the previous transaction date)
- Transaction Amount

# Summary

ROC AUC	Precision	Recall	Type I Error	Type II Error
0.9362	0.21	0.85	0.112	0.005



Q&A