

# 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

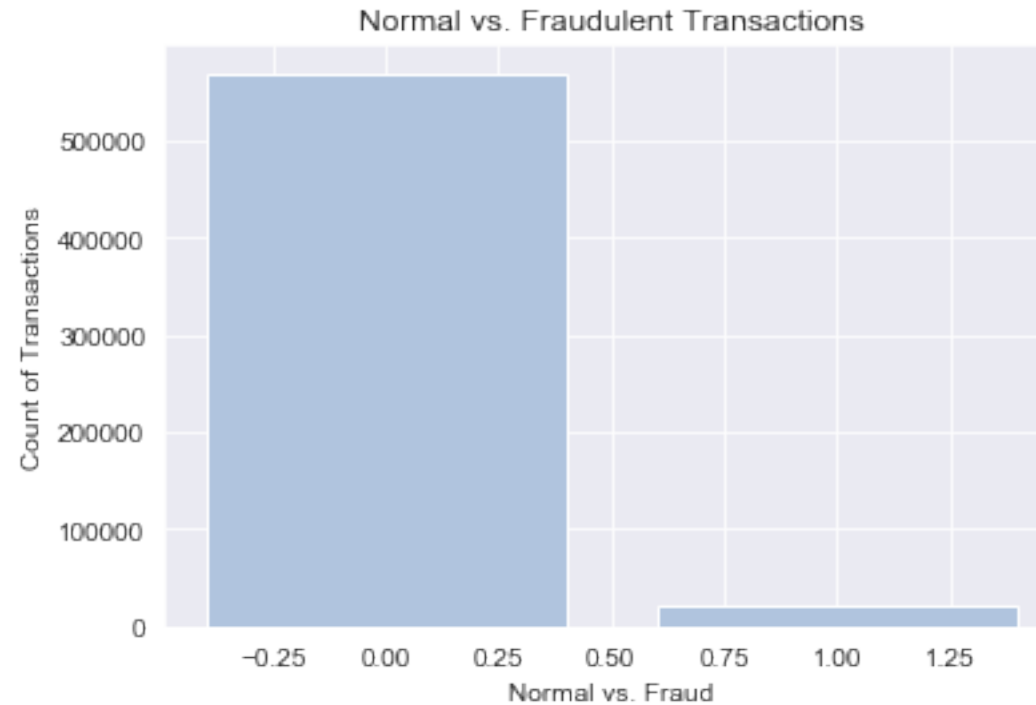
## I. Data Preparation:

- Data Cleaning
- Exploratory Data Analysis
- Feature Engineering

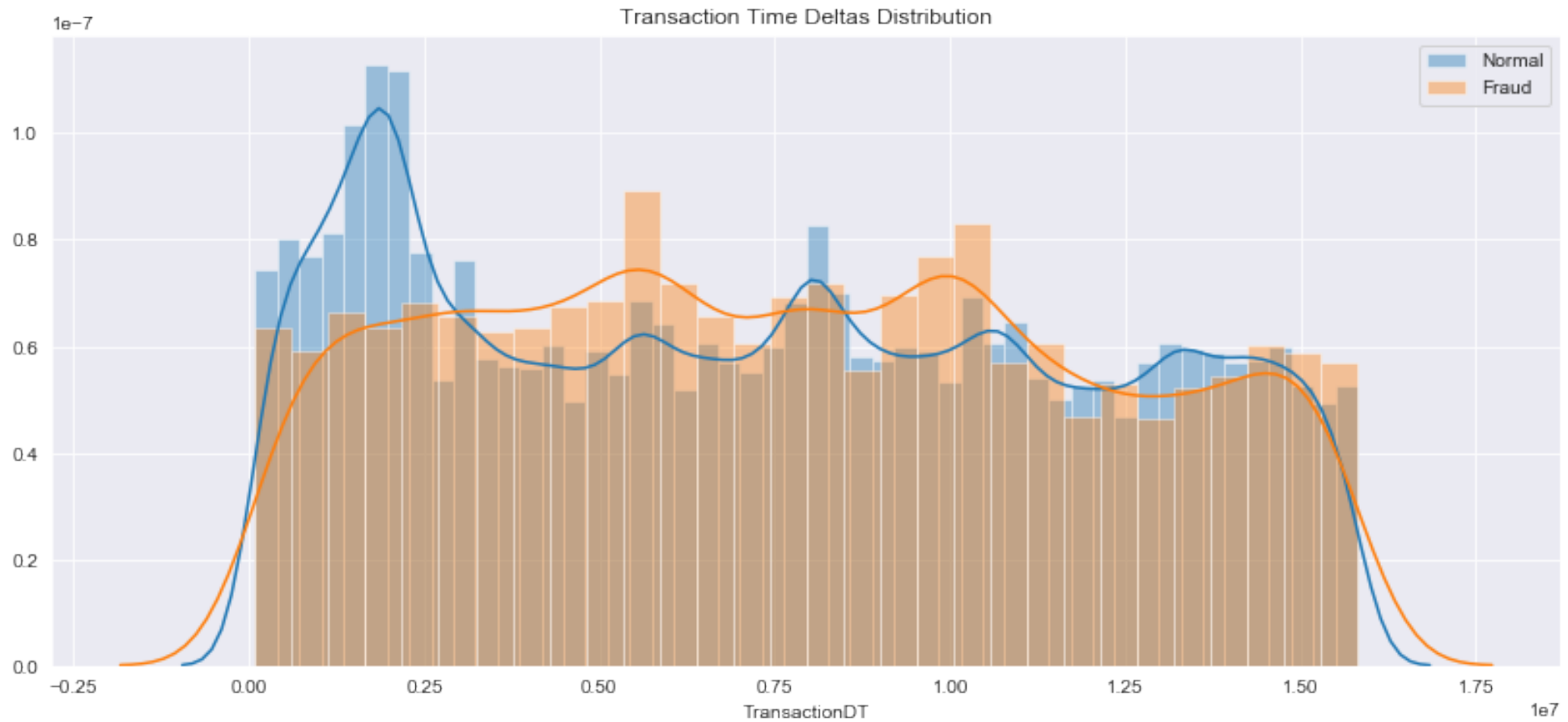
## II. Modeling:

- Class Imbalance
- Model Evaluation and Selection
- Hyperparameter Tuning

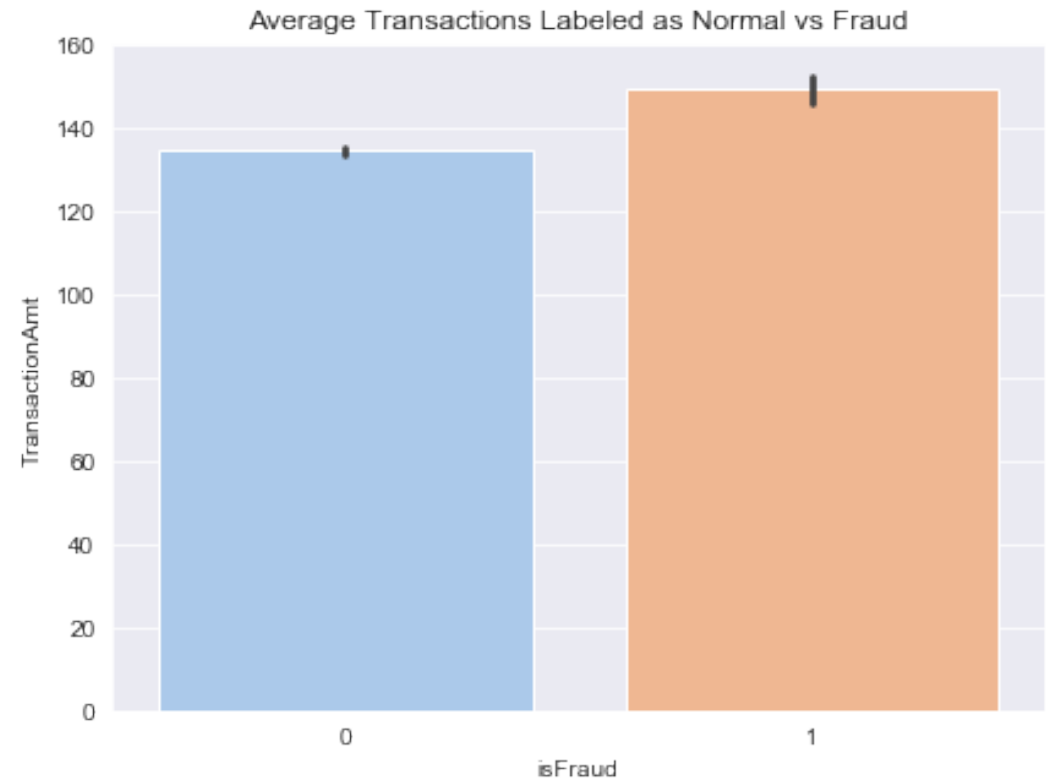
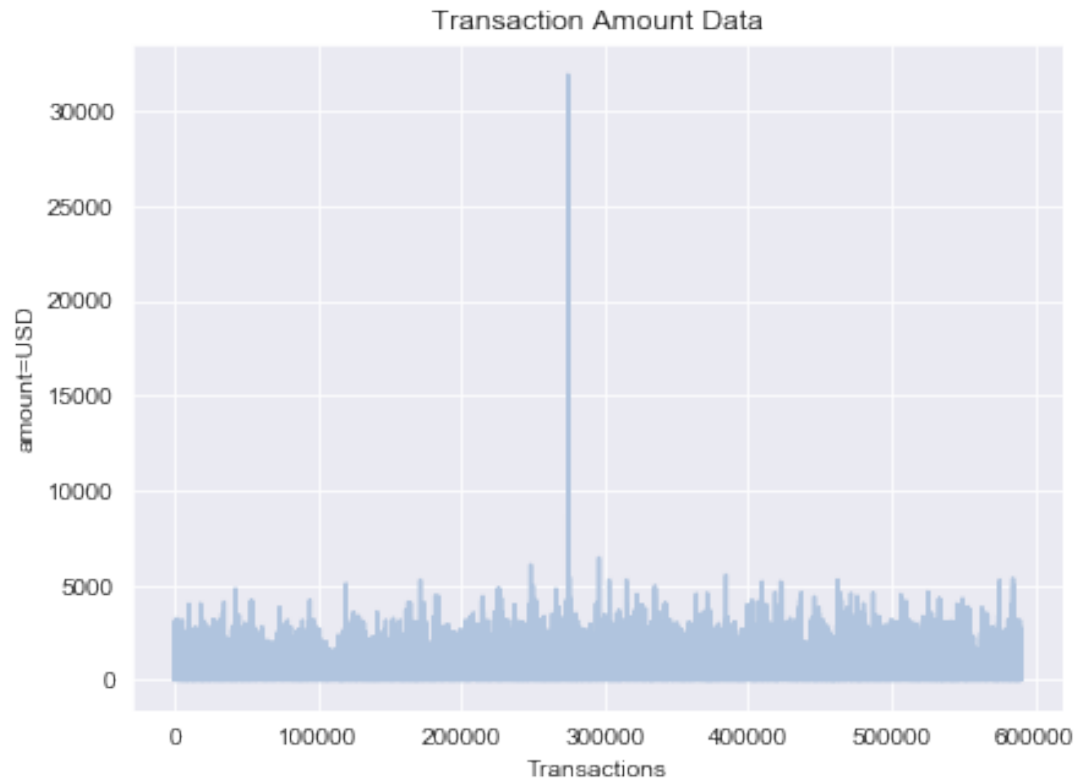
# Target Variable: class imbalance



# Transaction Time Deltas

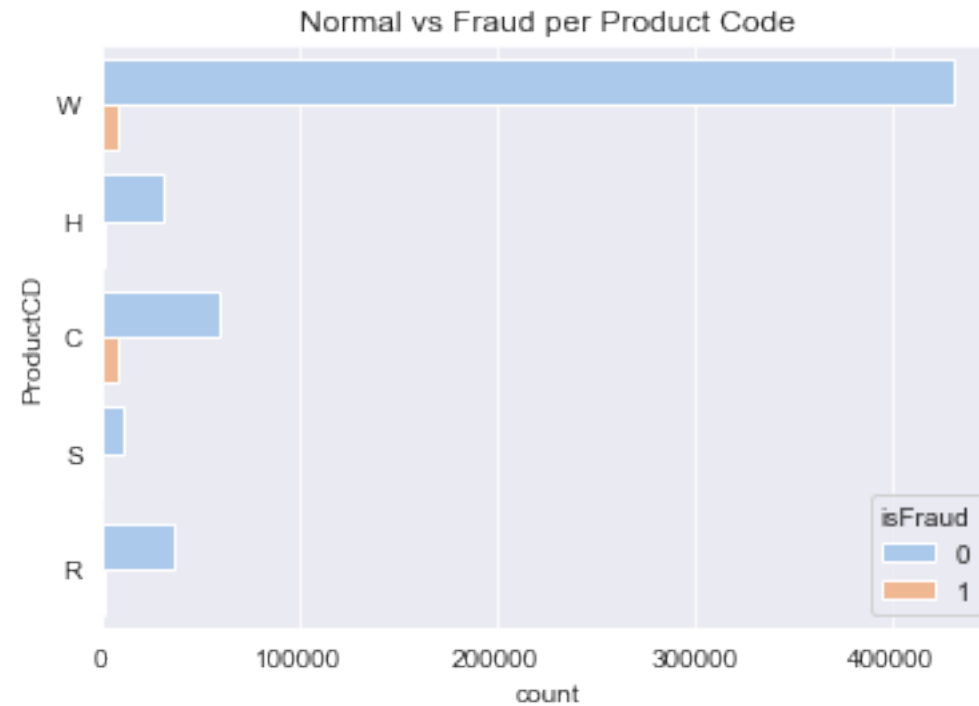


# Transaction Amount

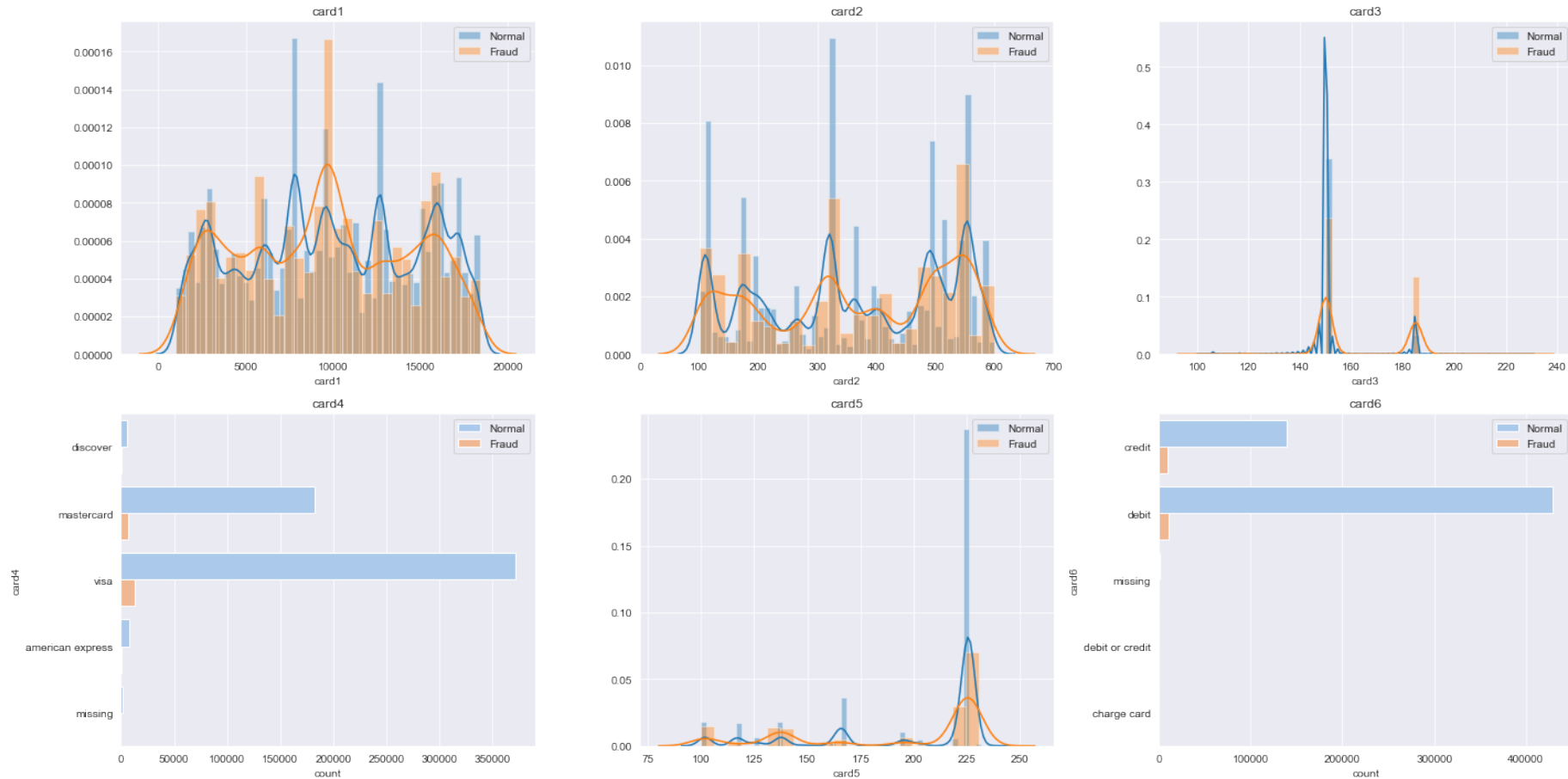




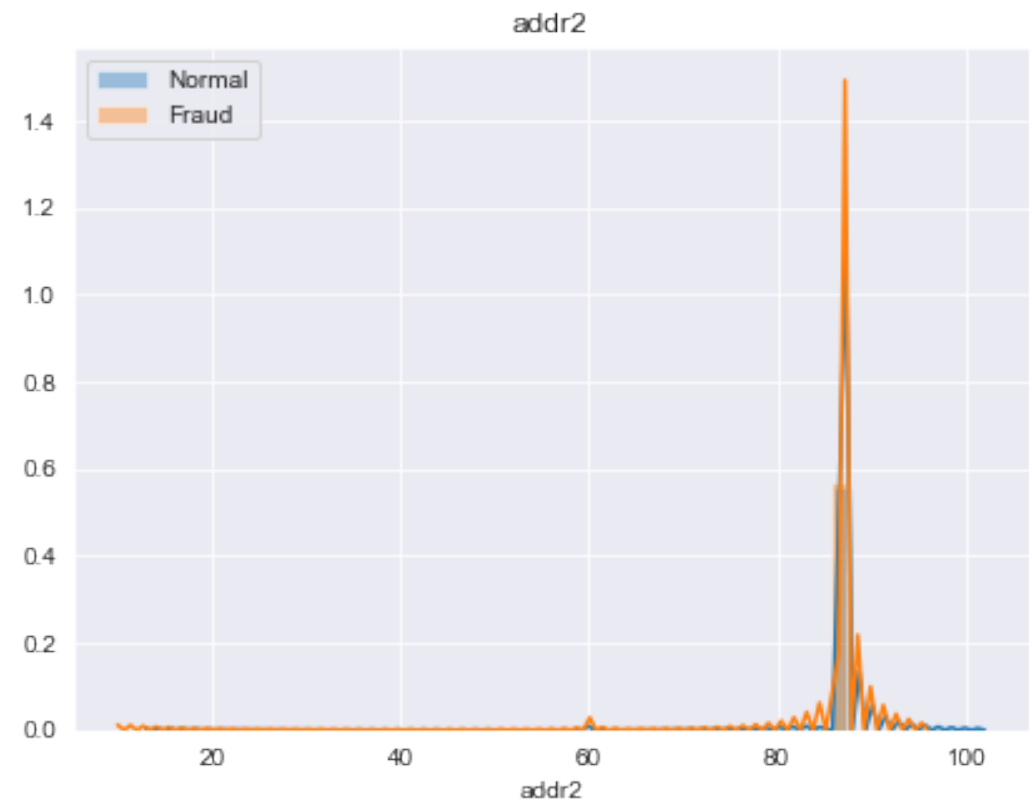
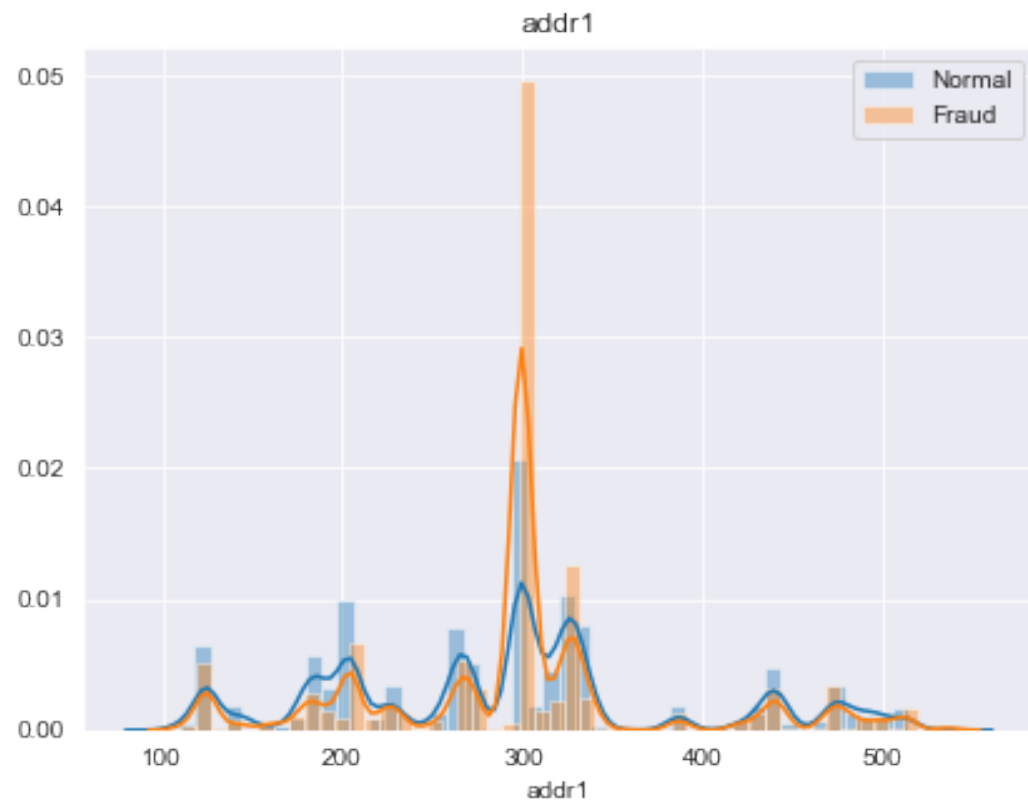
# Product Codes



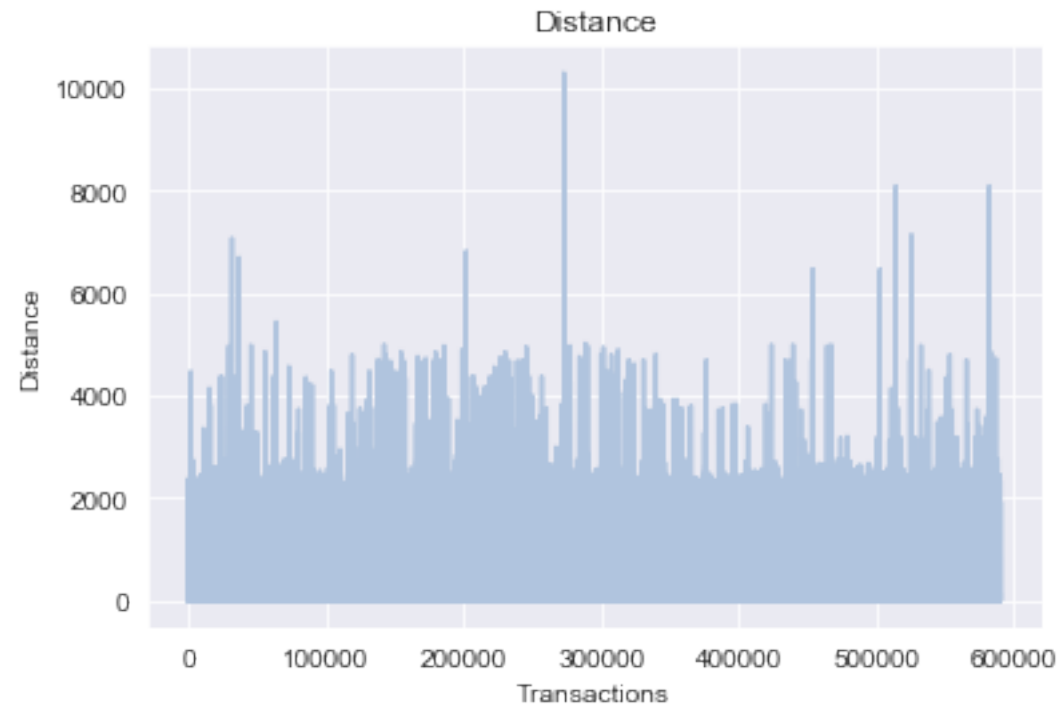
# Payment Card Information



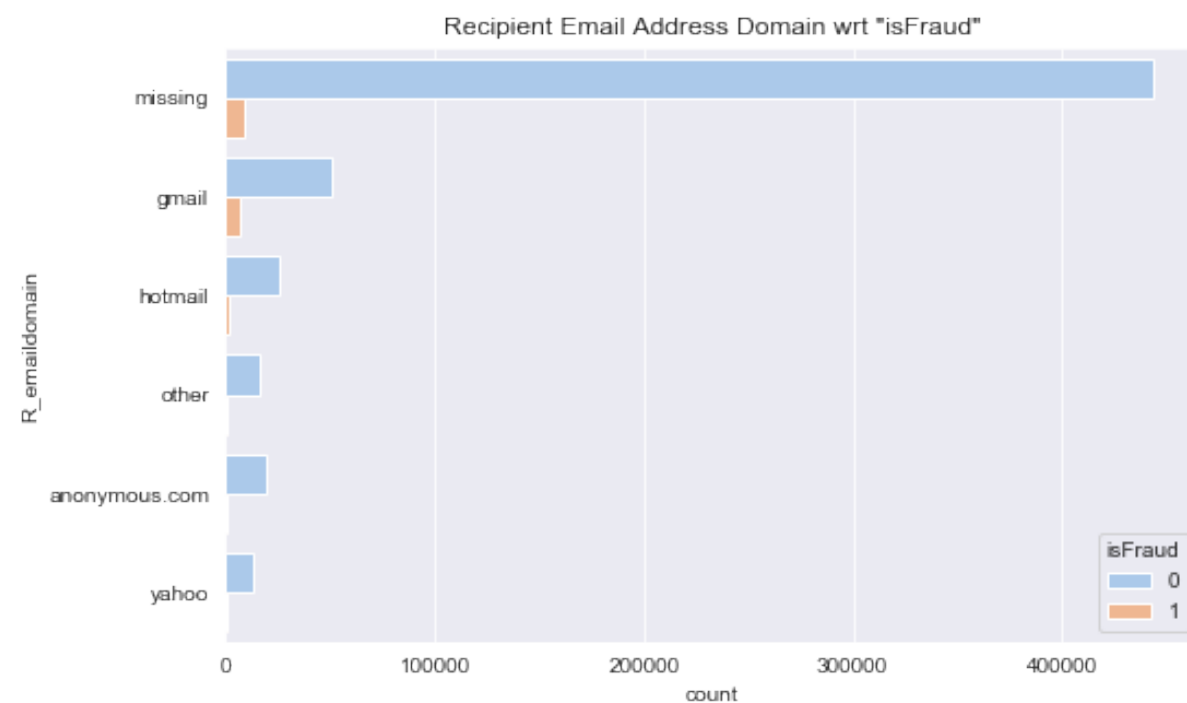
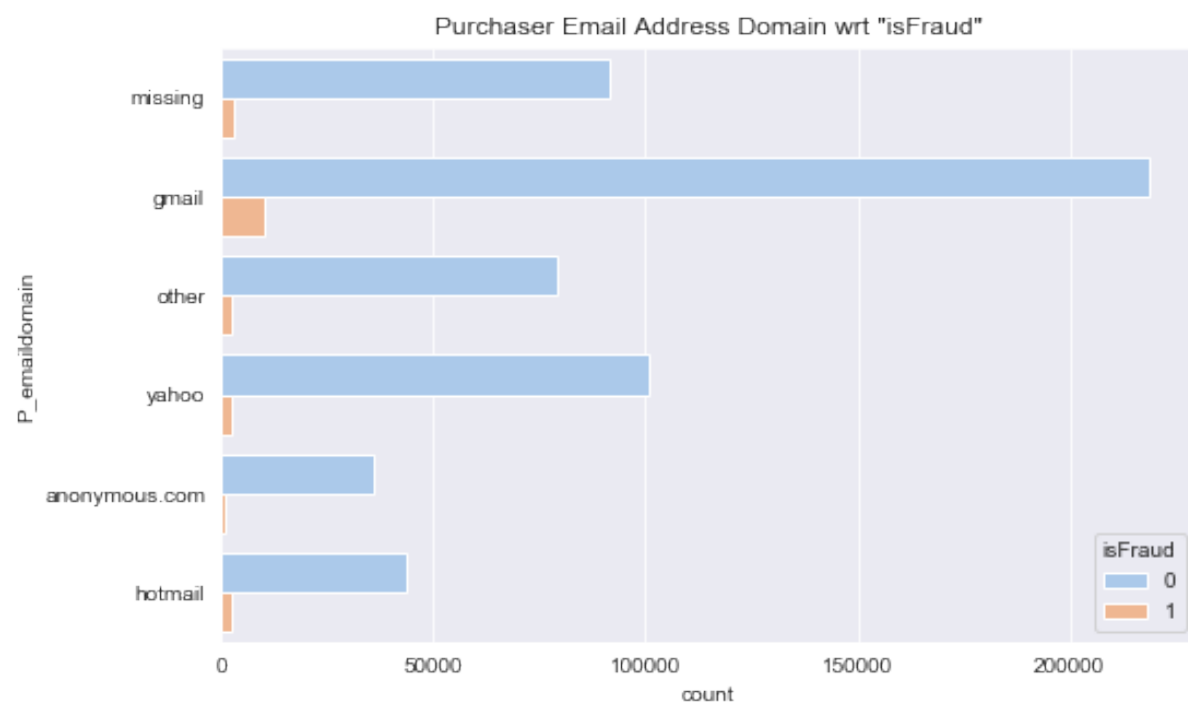
# Address



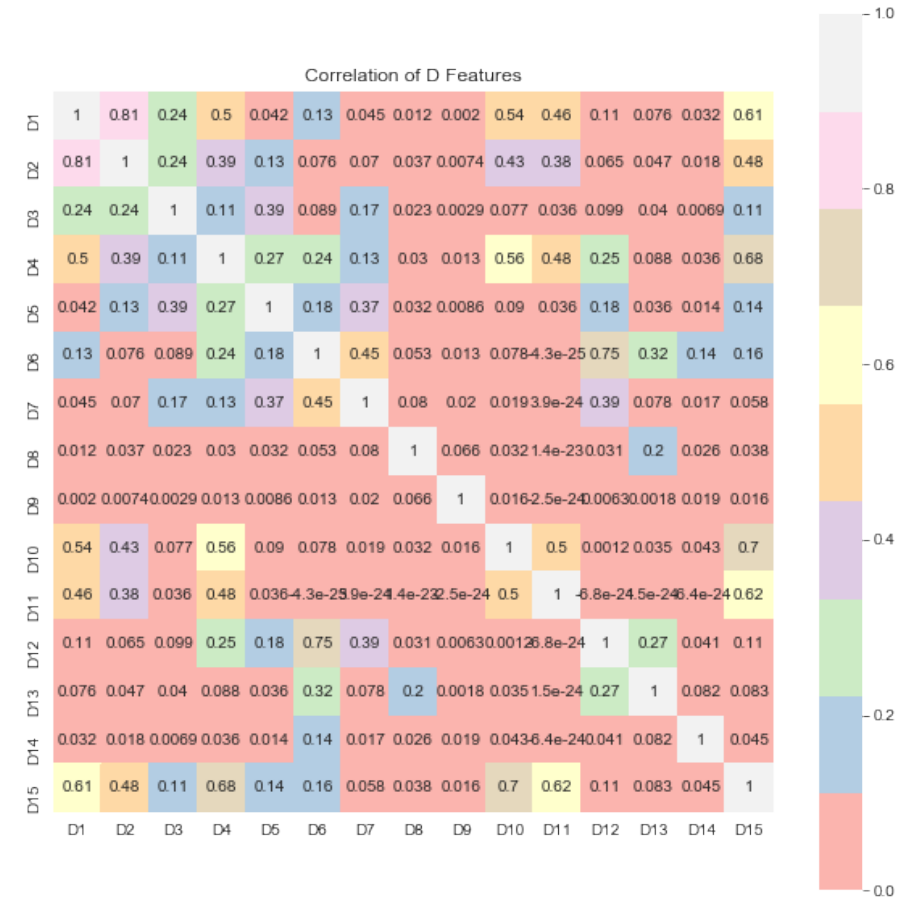
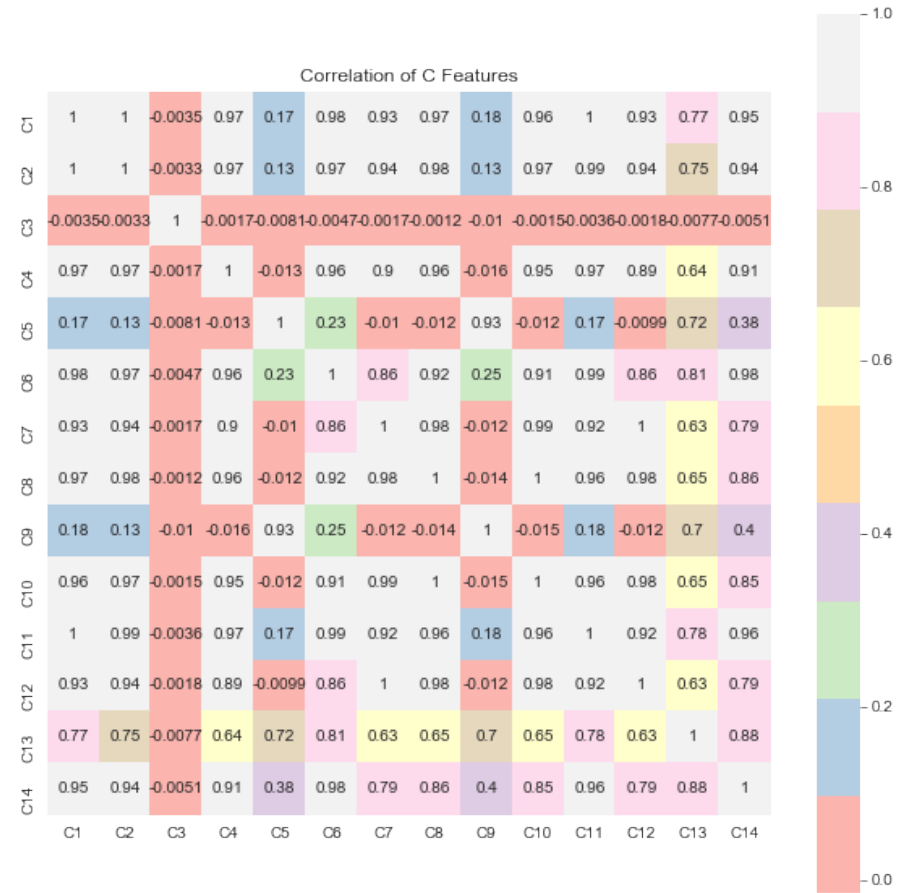
# Distance



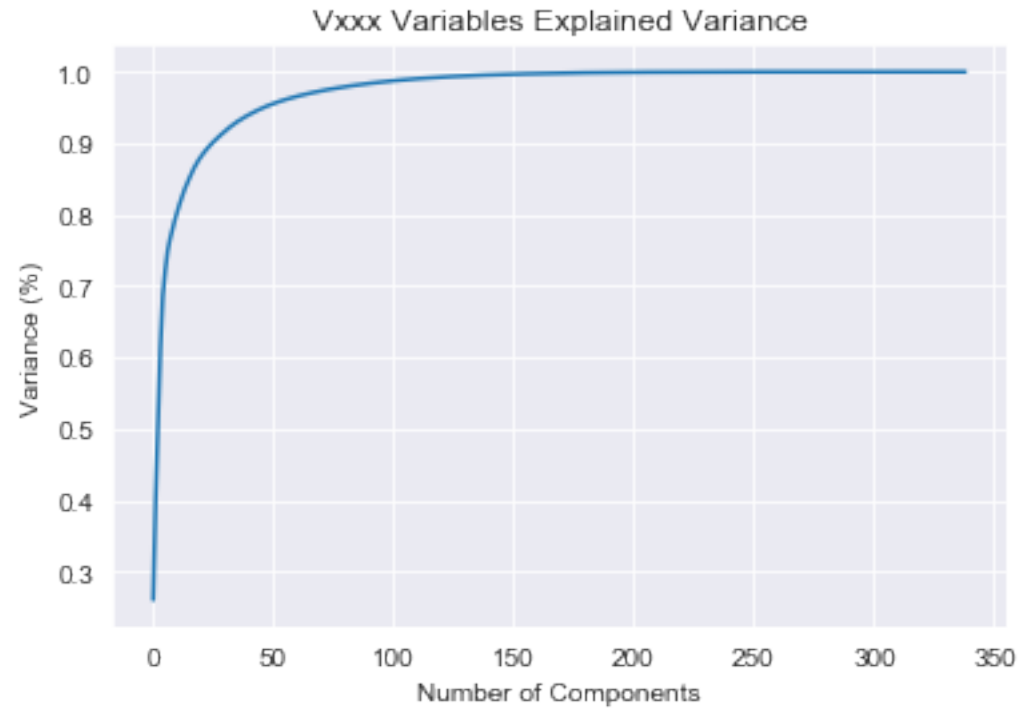
# Email Domains



# Feature Correlation

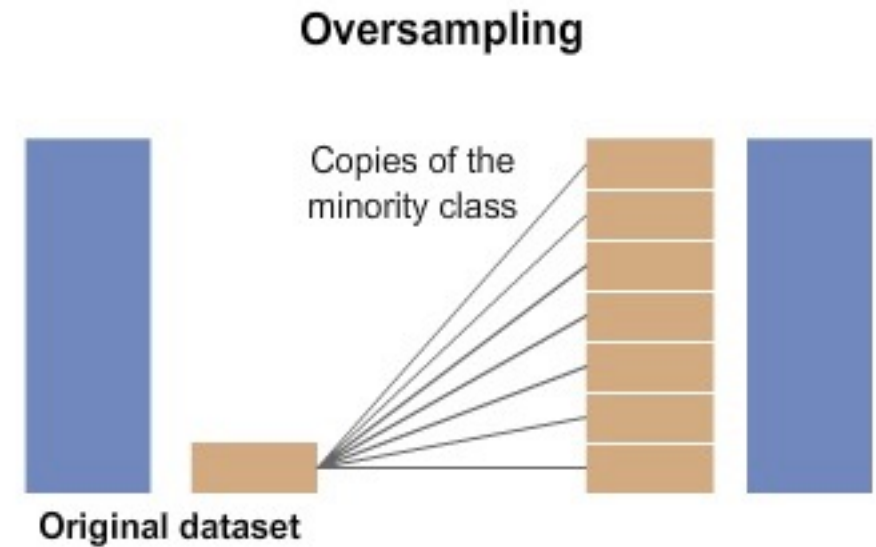
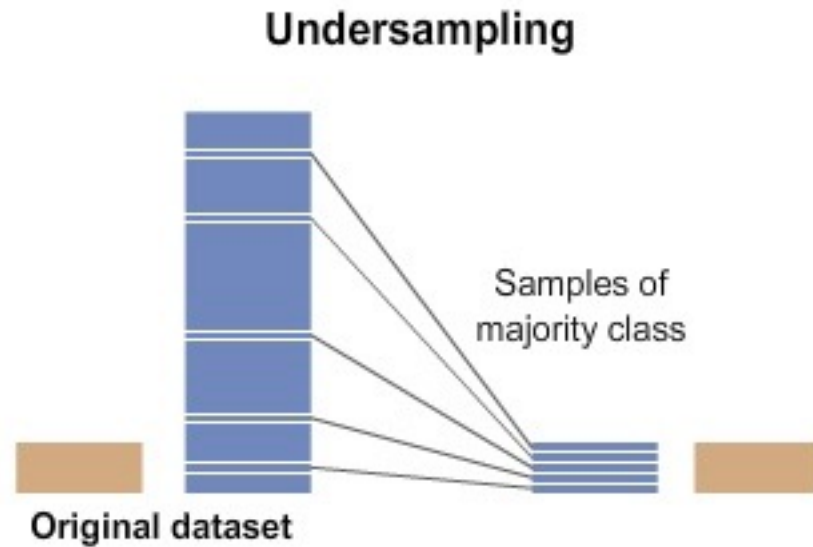


# Principal Component Analysis



# Class Imbalance

- Random Undersampling
- Synthetic Majority Oversampling Technique (SMOTE)

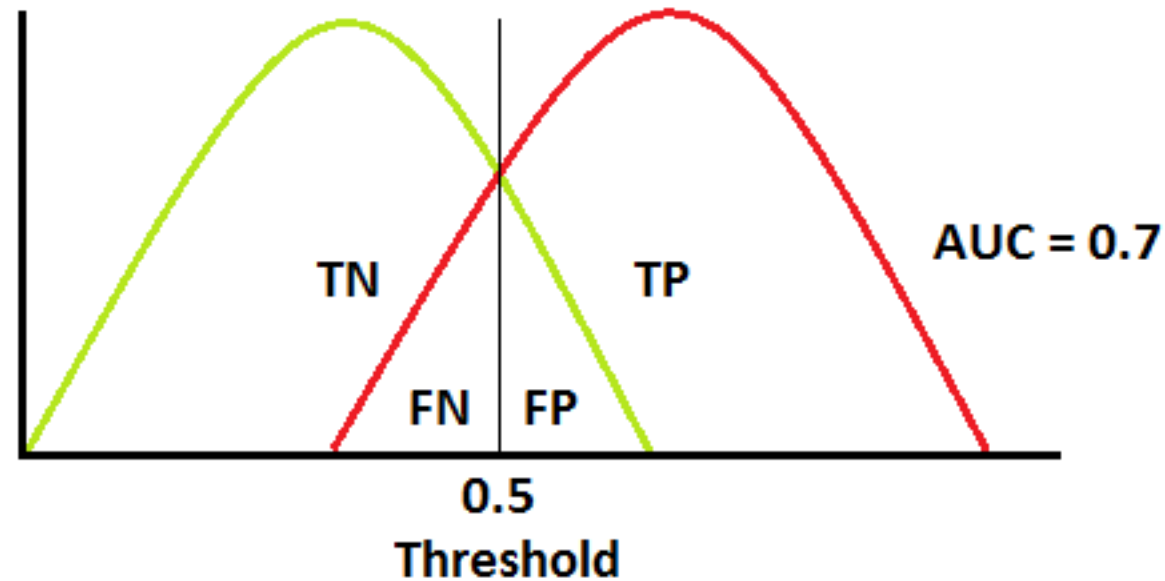


([source](#))



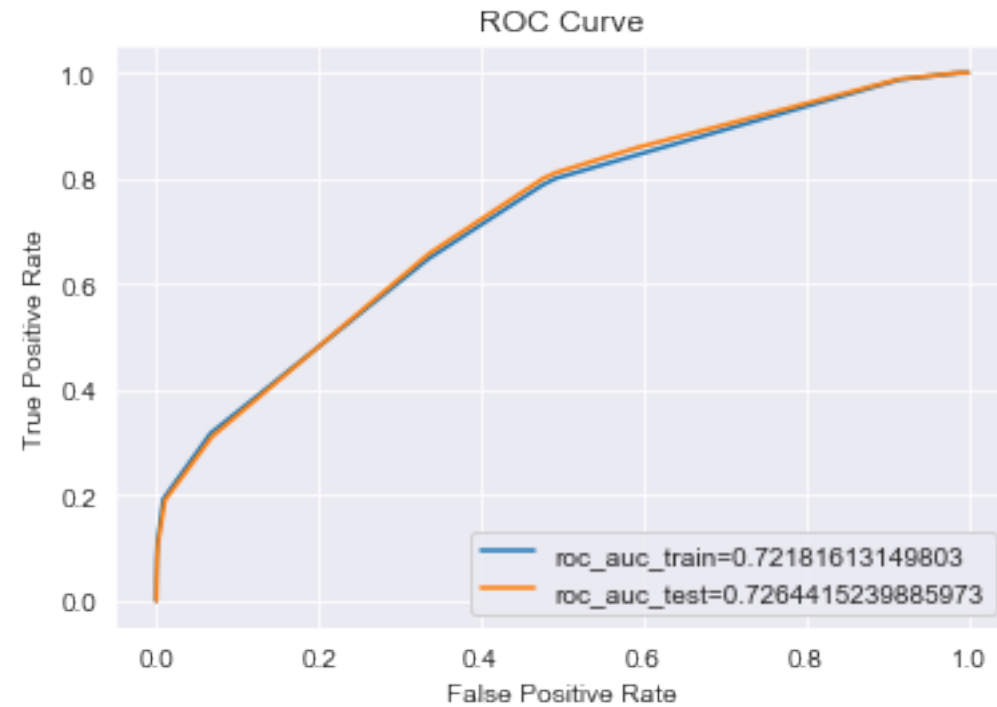
# Evaluation Metrics

- Area under Receiver Operating Characteristic curve
- 0.93

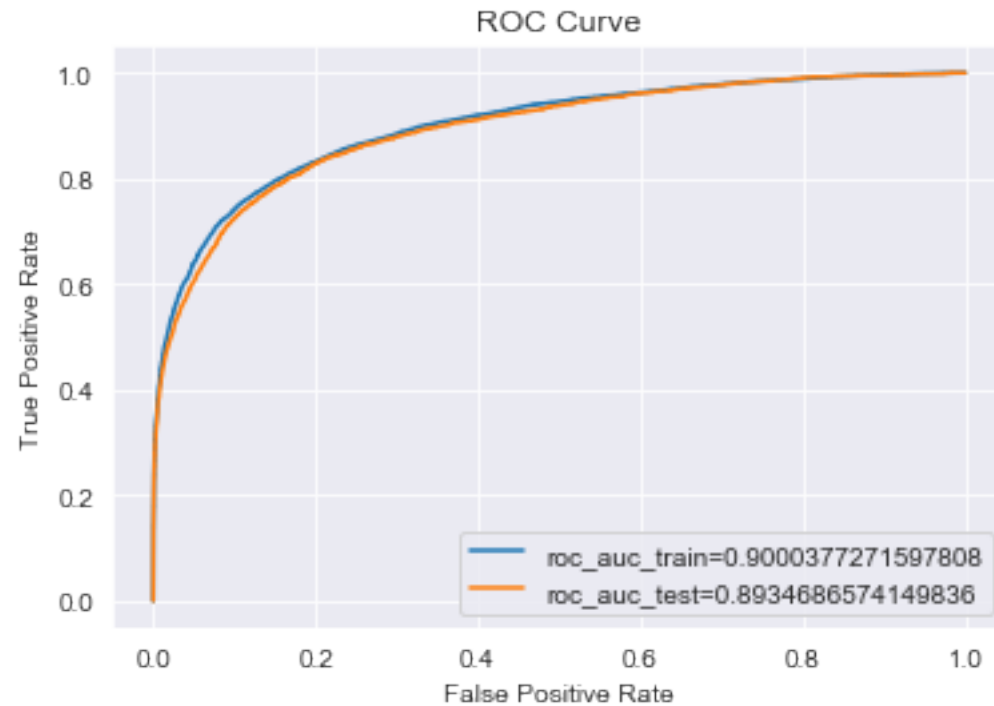


([source](#))

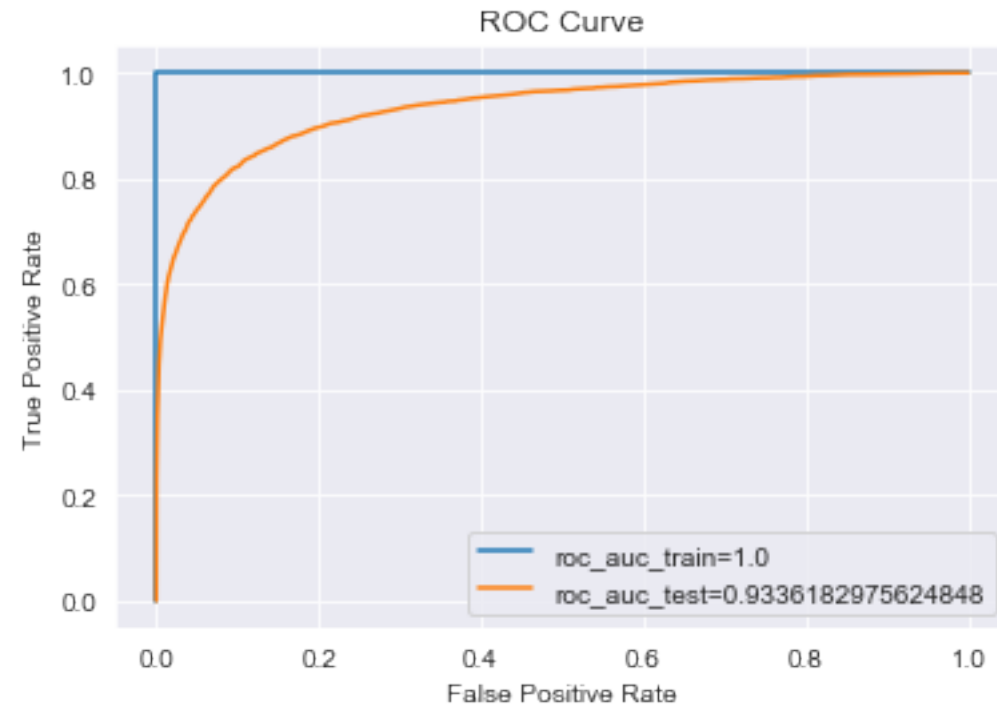
# Decision Tree



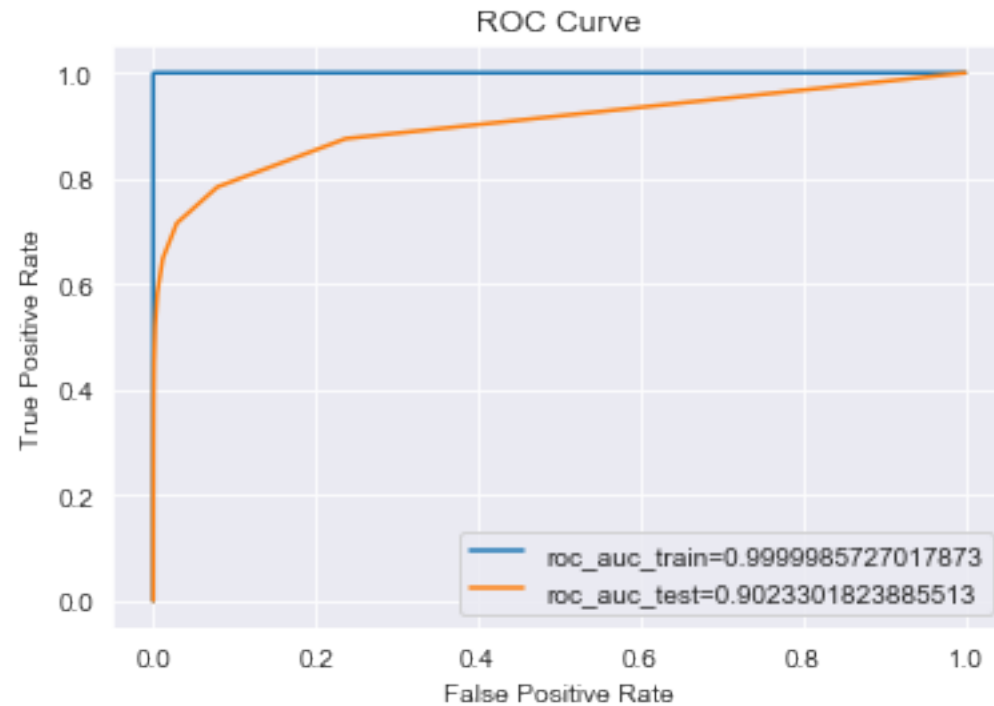
# Gradient Boosting Classifier



# Random Forest



# Random Forest - SMOTE



# Feature Importance ( $\geq 0.03$ )

- Transaction Time Deltas
- Transaction Amount
- C13
- I (V-feature)

# Summary

Model	AUROC	Precision	Recall	Type I Error	Type II Error
Random Forest (500)	0.9336	0.21	0.84	0.115	0.005
Random Forest (SMOTE)	0.9023	0.89	0.52	0.002	0.017

Q&A