A SERIES OF FORTUNATE FRAUD DETECTION

Fraud Detection :
Fraudsters and Phishers and Spammers...Oh, my!

Designing a Predictive Model to Catch E-commerce Bad Guys
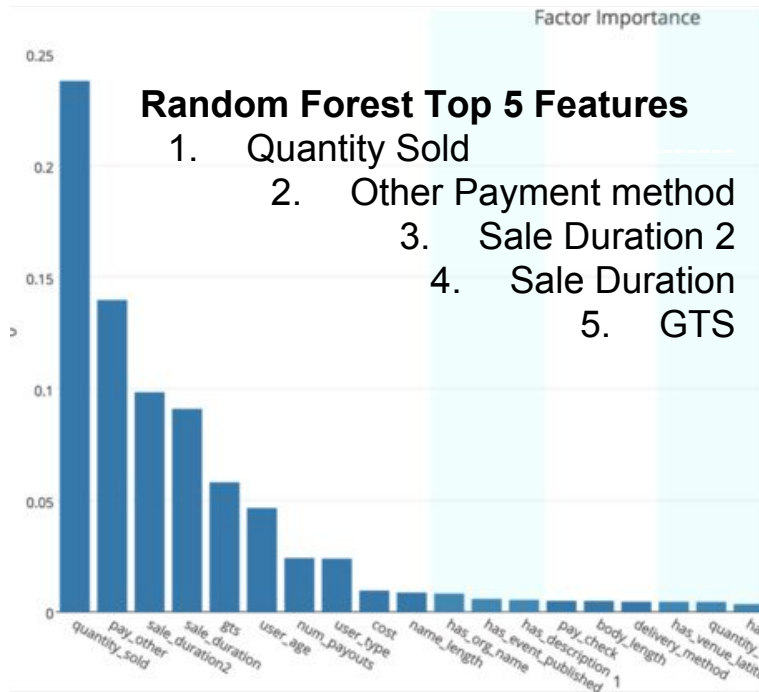
Exploring the Data

- **Data**
  - Lots of text
  - Variety of null data values
- **Feature Engineering**
  - NaN removal
  - Has_indicator feature (1/0)
    - Ex: has_twitter
  - Natural Language Processir

Labeled FRAUD by Location

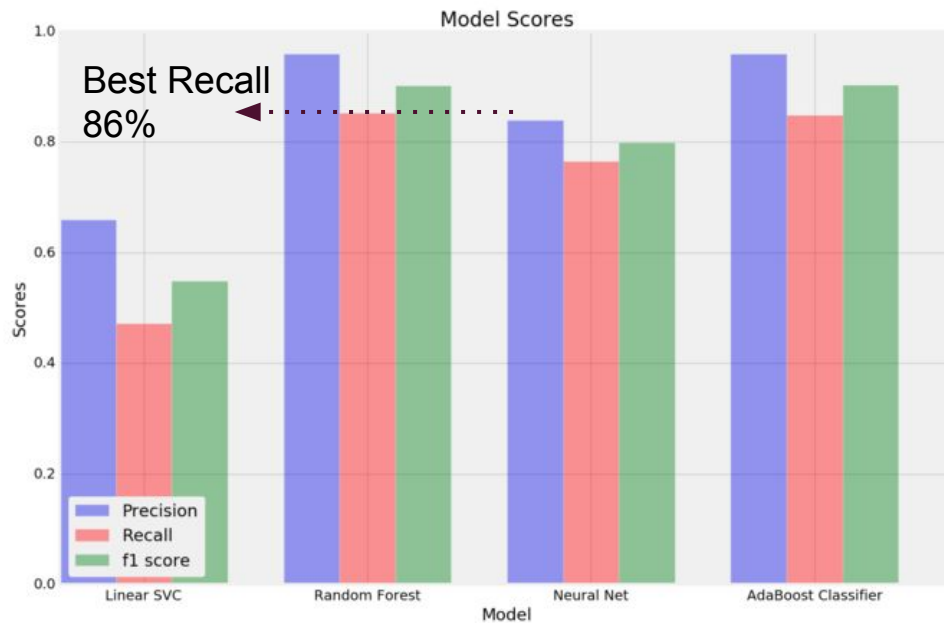Larger dot indicates a large financial loss due to fraud

Project Scope and Deliverables

- **PROJECT SCOPE**
    - To accurately identify potential fraud cases using machine learning.
- **DELIVERABLES**
    - Deploy a web app to predict fraud cases for further investigation as data comes in live.
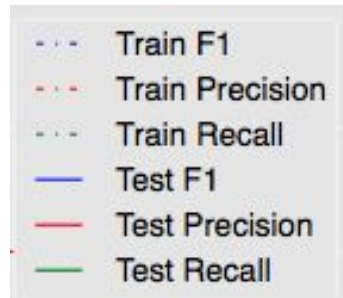   - http://ec2-52-86-7-15.compute-1.amazonaws.com:8106/

Random Forest Model – Features & Comparison

Factor Importance

Random Forest Top 5 Features
1. Quantity Sold
2. Other Payment method
3. Sale Duration 2
4. Sale Duration
5. GTS

Model Scores

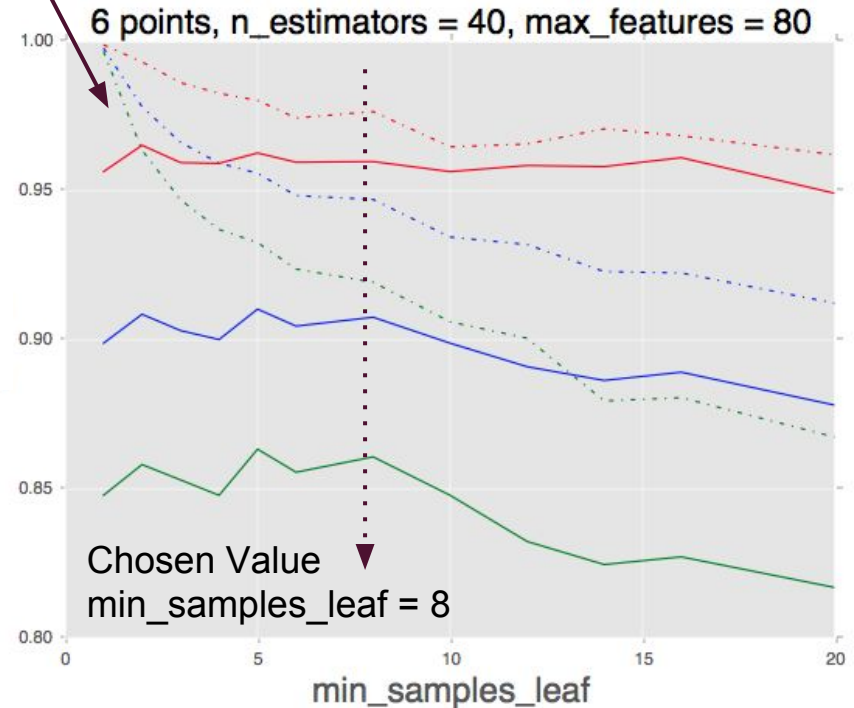Best Recall 86%

Linear SVC | **Random Forest** | Neural Net | Adaboost

Random Forest Hyper- Parameters

6 points, n_estimators = 40

Score

max_features

Overfitting

Chosen Value
max_features = 80
(643 total)

No Overfitting

6 points, n_estimators = 40, max_features = 80

Chosen Value
min_samples_leaf = 8

min_samples_leaf

- - - Train F1
- - - Train Precision
- - - Train Recall
——— Test F1
——— Test Precision
——— Test Recall

Top Words Associated With

Not Fraud

Fraud

Top words came from Natural Language Processing from event and organization descriptions
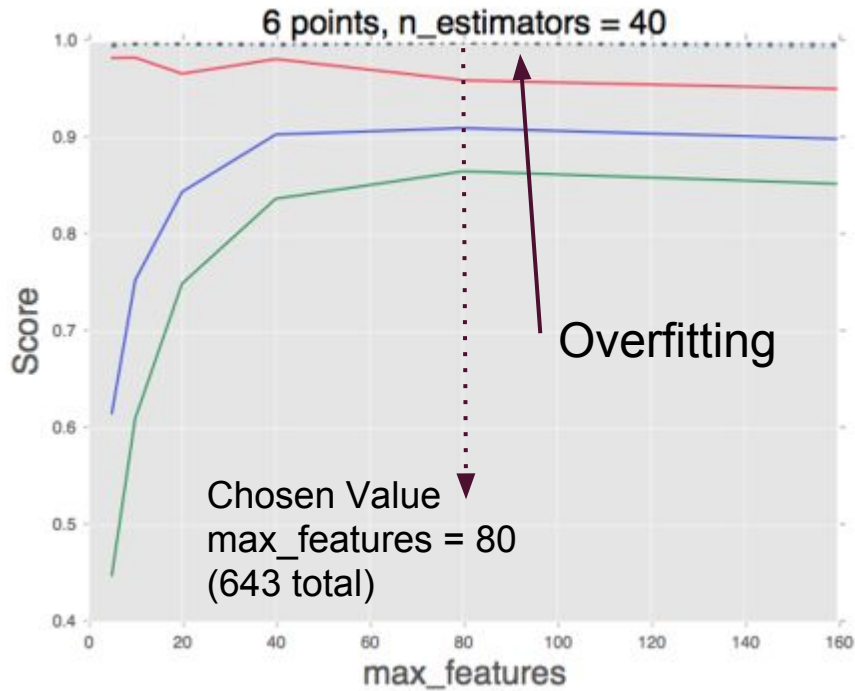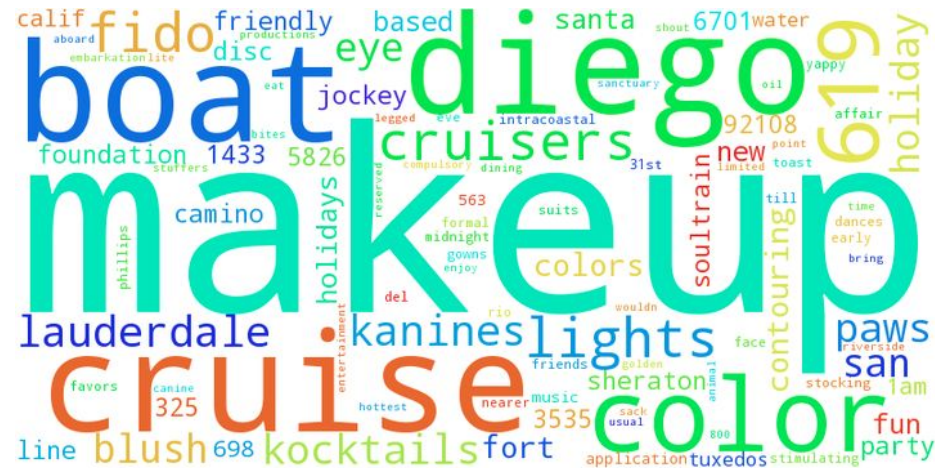
Cost Benefit Matrix

& Profit Curve

	Fraud	Not Fraud
Fraud	-$50	$50
Not Fraud	-$500	$100

Profit Curve

— Random Forest

The average # of fraud attempts against the company constitute about 9% of our transactions. We need to be able to capture ½ of these cases to maximize our profits.

Profit

Percentage of instances (decreasing by score)
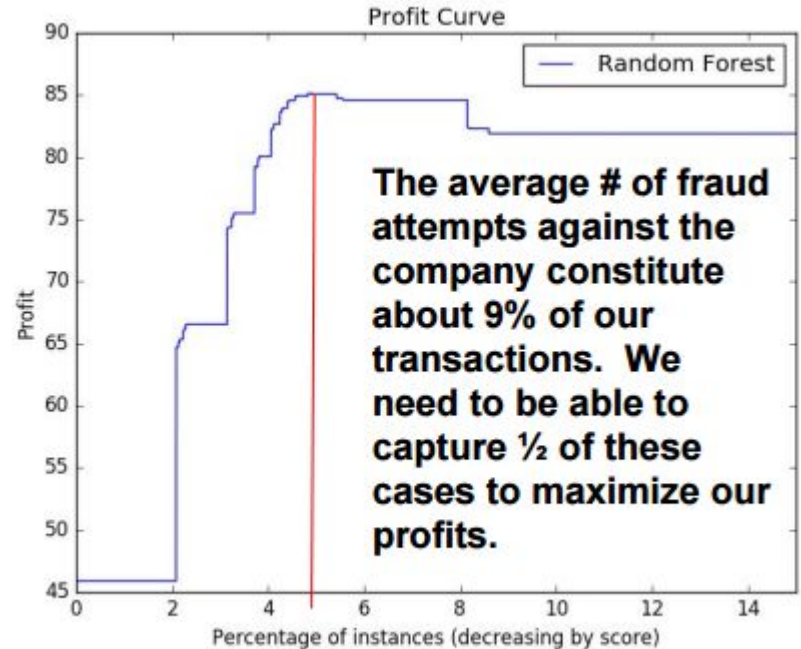
ROC Curve
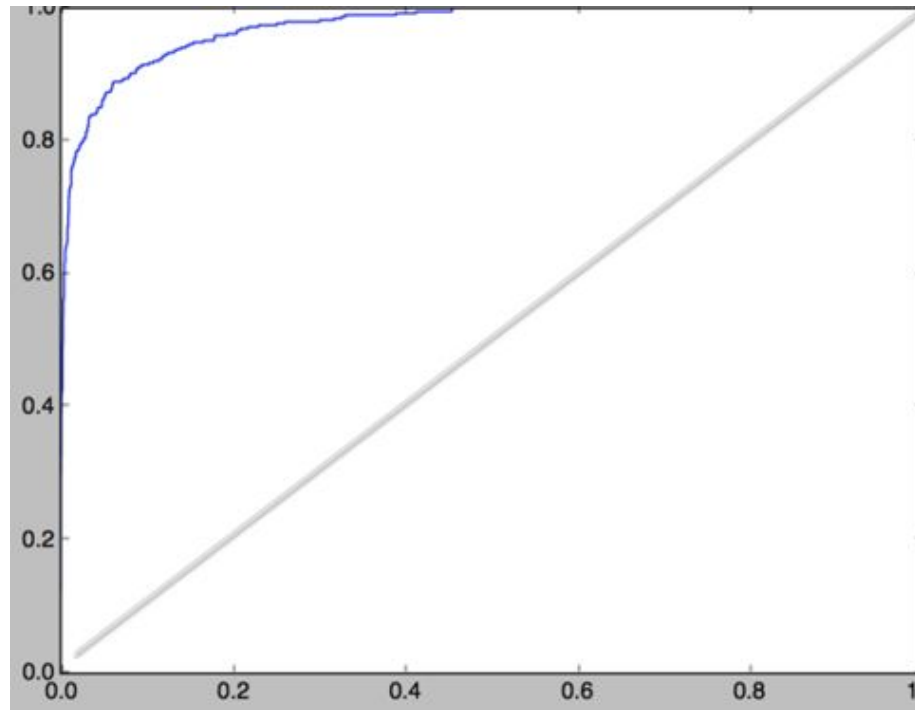
Prediction Highlights

Book Now For Our 2 Hour Seminar

- How to replace your income in as little as 2 years
- How to pay your home off in 6 years
- Understand market timing: learn about the property clock and wh
- How to reduce your taxes legally, potentially to £0
- Build a balanced real estate portfolio structured for po
- How to diversify your portfolio across different market
- Learn the secrets ONLY 5% of investors know that gen
- Discover how YOU can find properties that pay you tho
- Learn how to add massive value, create instant equity
- Avoid common finance traps that destroy the investing
- Learn effective debt reduction secrets that most peopl
- How to save time by having a team of experts source

Prediction: Potential FRAUD

Prediction: Not fraud

Description:

Thanks for being awesome

Prediction: Potential FRAUD

Description:

You can get money from our gameshow. In a lot of people who joined in our game show, u must choose our question asap. The computer will choose the fastest people who answer our question correct and faster more than all people. So, u have to 15 question multiple choice. U will get some money from ur correct question.

Labeled FRAUD by Location

Larger dot indicates a large financial loss due to fraud