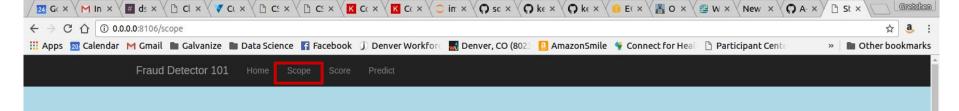# Fraud Detection :
Fraudsters and Phishers and Spammers...Oh, my!

Designing a Predictive Model to Catch E-commerce Bad Guys

Project Scope

Goal

To accurately identify potential fraud cases for further investigation using machine learning.

A created model will then be hosted on an app to detect fraud cases live.

Deliverables

Deployed web app to predict fraud cases for further investigation as data comes in live.

Workflow

Exploratory Data Analysis:

Create training fraud labels

Remove NANs
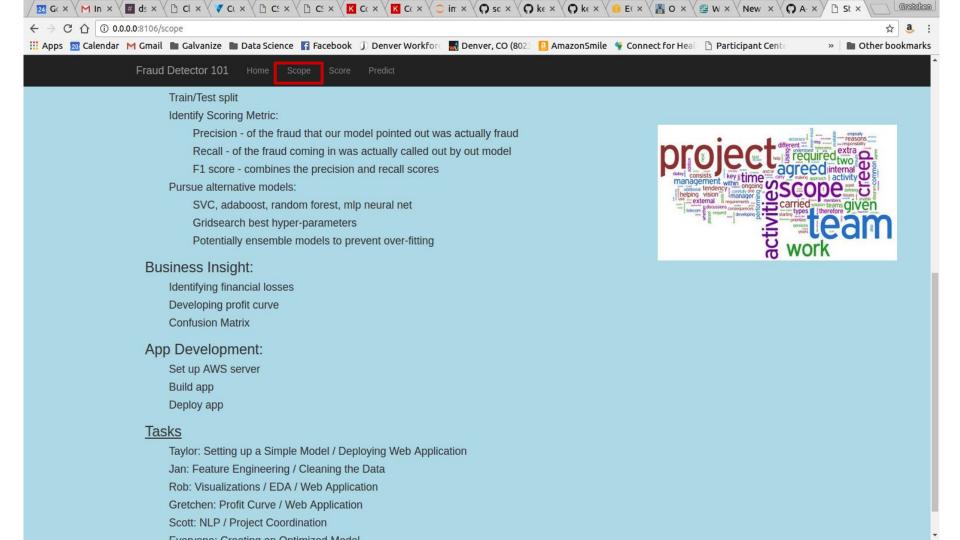
Plot raw data

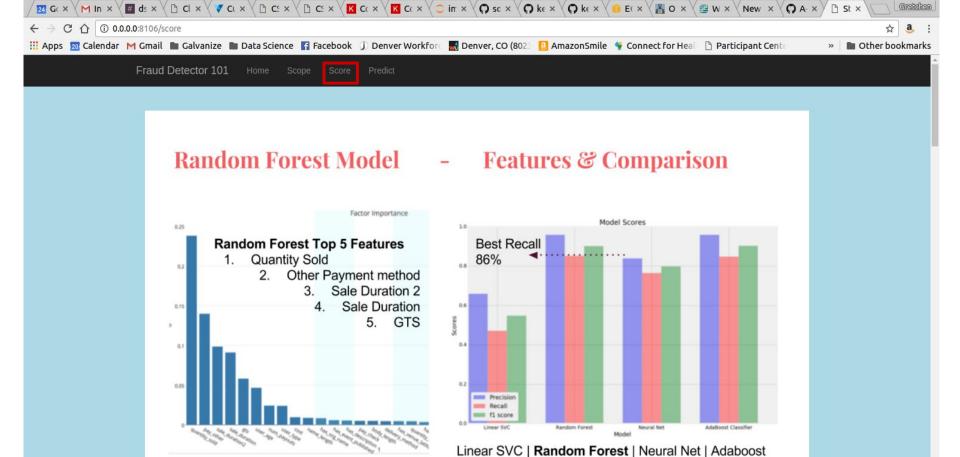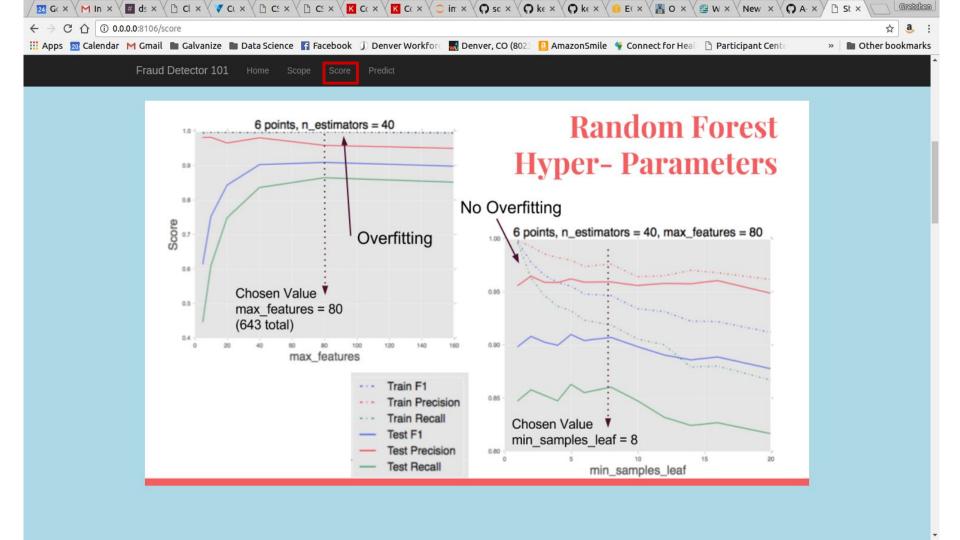Engineer features

Implement NLP analysis

Dropping features

Modeling:

Create baseline working model

Train/Test split

Identify Scoring Metric:

Train/Test split

Identify Scoring Metric:

    Precision - of the fraud that our model pointed out was actually fraud

    Recall - of the fraud coming in was actually called out by out model

    F1 score - combines the precision and recall scores

Pursue alternative models:

    SVC, adaboost, random forest, mlp neural net

    Gridsearch best hyper-parameters

    Potentially ensemble models to prevent over-fitting

Business Insight:

    Identifying financial losses

    Developing profit curve

    Confusion Matrix

App Development:

    Set up AWS server

    Build app

    Deploy app

Tasks

    Taylor: Setting up a Simple Model / Deploying Web Application

    Jan: Feature Engineering / Cleaning the Data

    Rob: Visualizations / EDA / Web Application

    Gretchen: Profit Curve / Web Application

    Scott: NLP / Project Coordination

Random Forest Model   -   Features & Comparison

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

Random Forest Hyper- Parameters

6 points, n_estimators = 40

Overfitting

Chosen Value
max_features = 80
(643 total)

No Overfitting

6 points, n_estimators = 40, max_features = 80

Chosen Value
min_samples_leaf = 8

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

ROC Curve

False Positive Rate

True Positive Rate

AdaBoost
RandomForest

Very high True Positive Rate &
Very low False Positive Rate

Top Words Associated With

Not Fraud

Fraud

Top words came from Natural Language Processing from
event and organization descriptions

Prediction Highlights

Prediction: Not fraud

Description:

Thanks for being awesome

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

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.

Prediction: Potential FRAUD
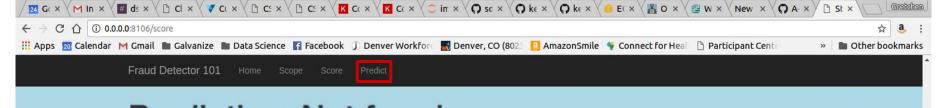
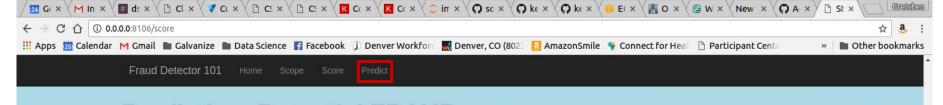# Prediction: Not fraud

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.