Predictive Modeling of Opioid Prescription Fraud by CMS Providers

DATA606 Capstone Project - Part 3
William Rubin – Spring 2020

PROJECT TIMELINE

Part 1
Problem Definition

Part 2
**Existing Research /
Initial EDA**

Part 3
**Final EDA /
Predictive Model Preparation**

Part 4
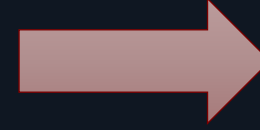**Execution, Interpretation,
and Results**

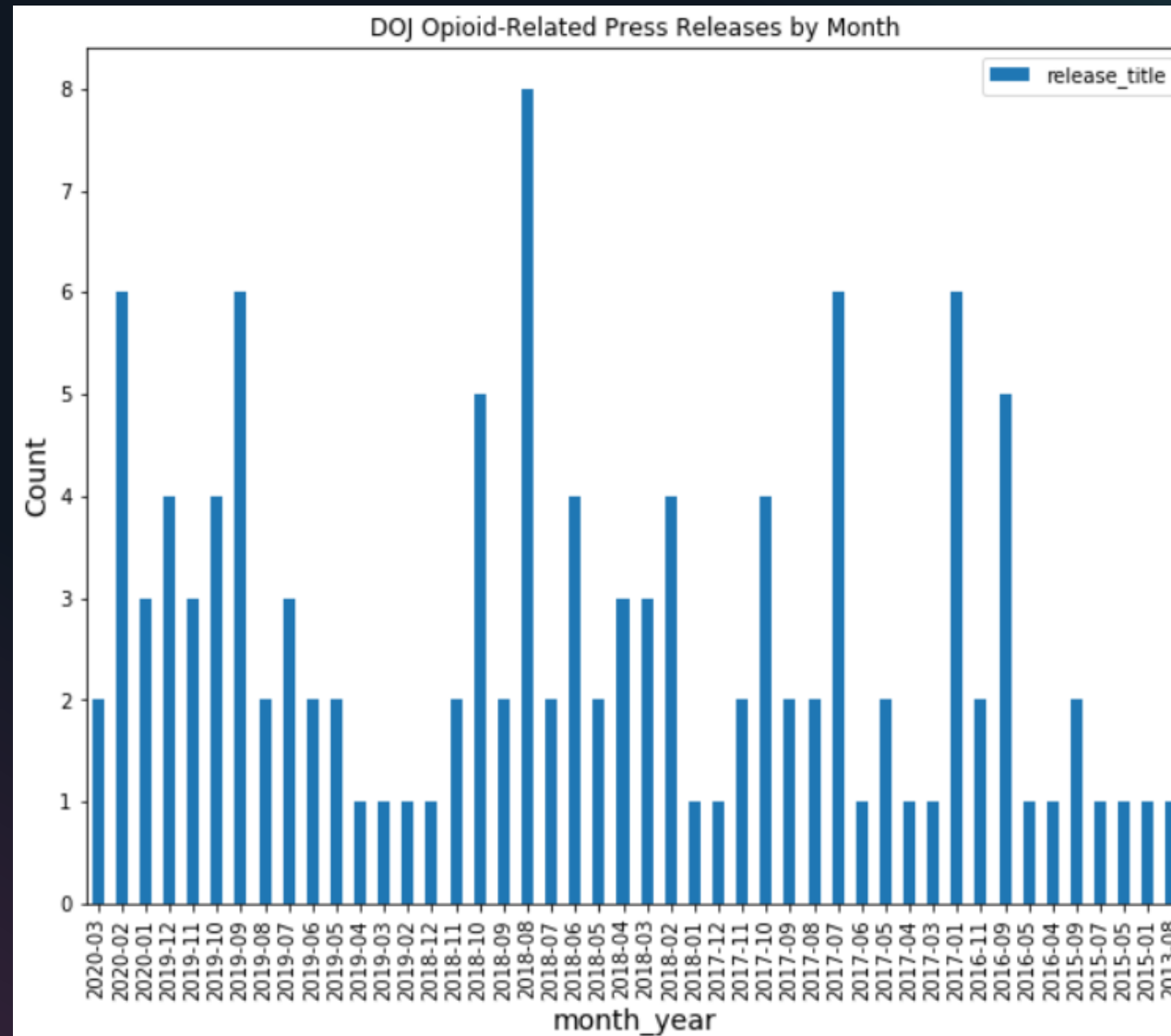# Part 3 Main Goal

Identify Fraudulent Providers

Determine Likely Legitimate Providers

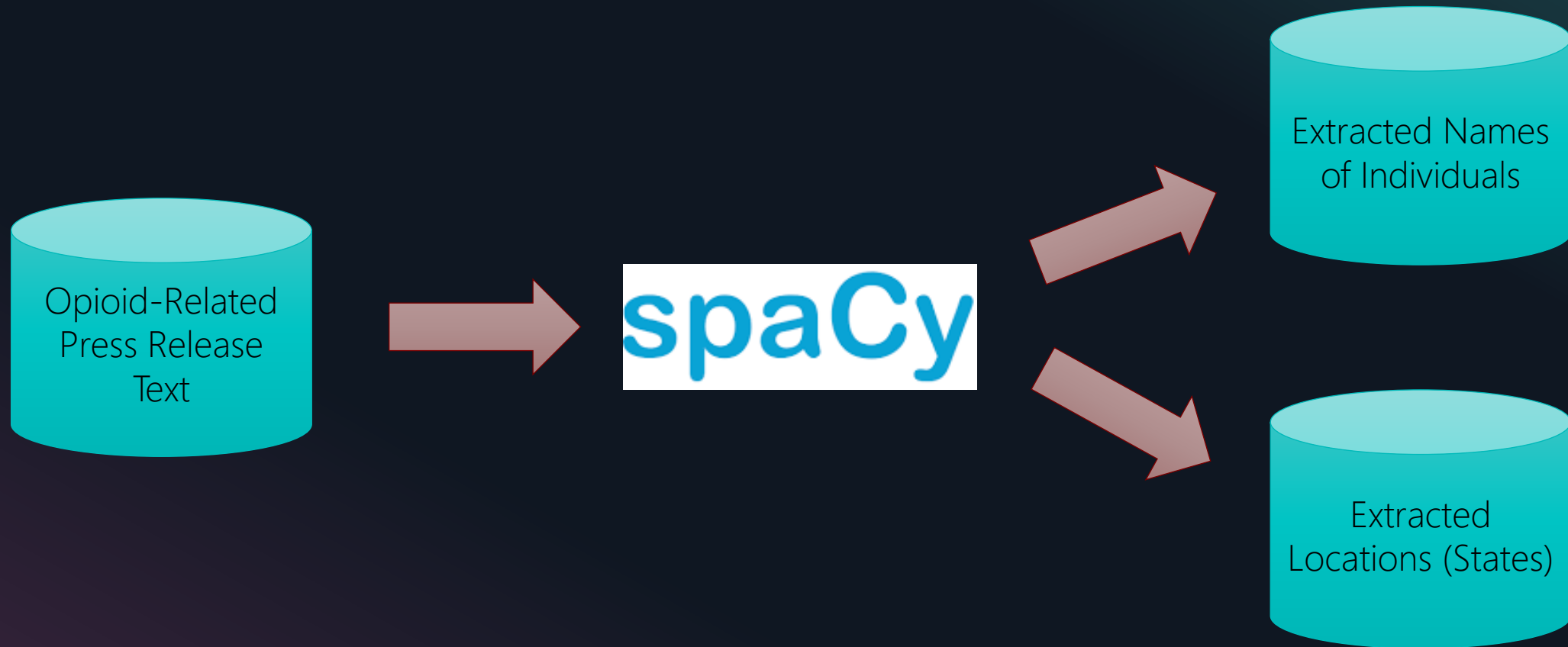*Create Labeled Opioid Provider Dataset*

Step 1: Data Collection

Opioid-Related
Press Release
Text

Step 1: Data Collection

DOJ Opioid-Related Press Releases by Month

Step 2: Entity / Location Extraction

Step 2: Entity / Location Extraction

• Sample Intermediate Output

Title of Press Release: "Four Detroit-Area Physicians Found Guilty of Health Care Fraud Charges for Role in Over $150 Million Health Care Fraud Scheme" for row 19
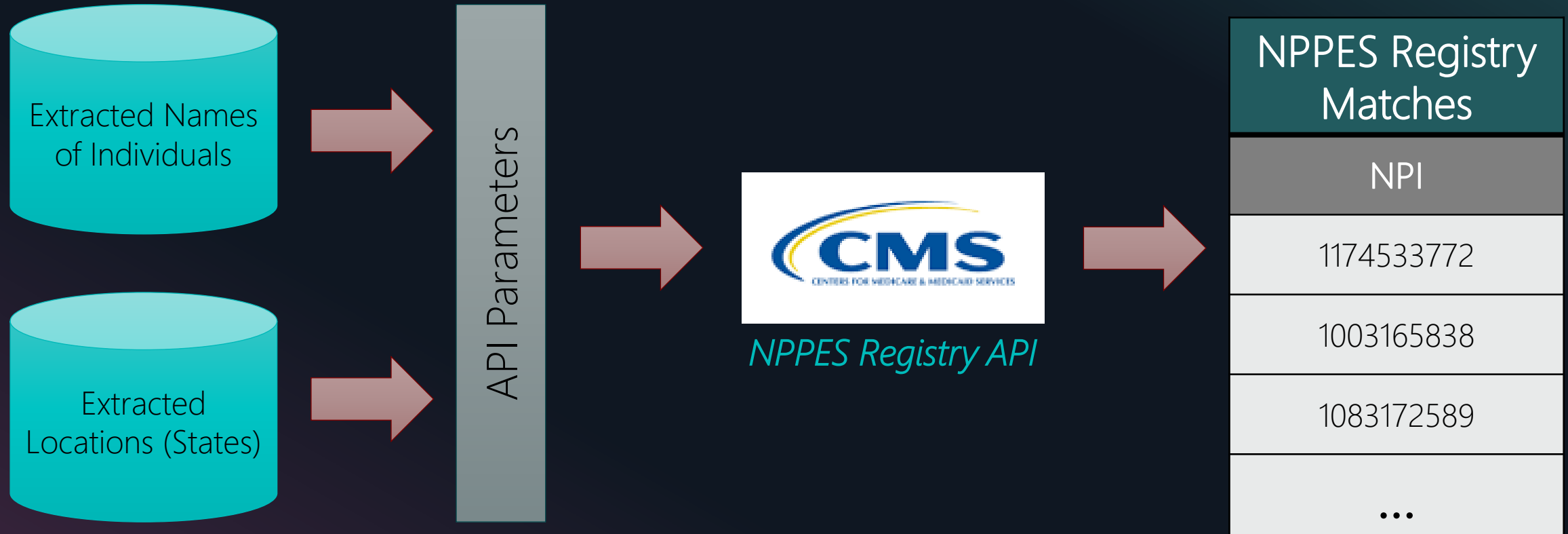**Number of Identified People**: 13
Number of Identified, Unique States: 2

Title of Press Release: "South Florida Doctor Convicted Of Participating in a Conspiracy to Illegally Dispense Opioids and Other Drugs" for row 63
Number of Identified People: 13
**Number of Identified, Unique States**: 1

Step 3: NPI Resolution

Extracted Names of Individuals

Extracted Locations (States)

API Parameters

NPPES Registry API

NPPES Registry Matches
NPI
1174533772
1003165838
1083172589
...

Step 3: NPI Resolution

Overview of Matches:

- 856 Overall Matches Found

- 260 Exact Matches Found

- Further Analysis Needed to Consolidate Records with Multiple Matches

NPPES Matching Results Histogram
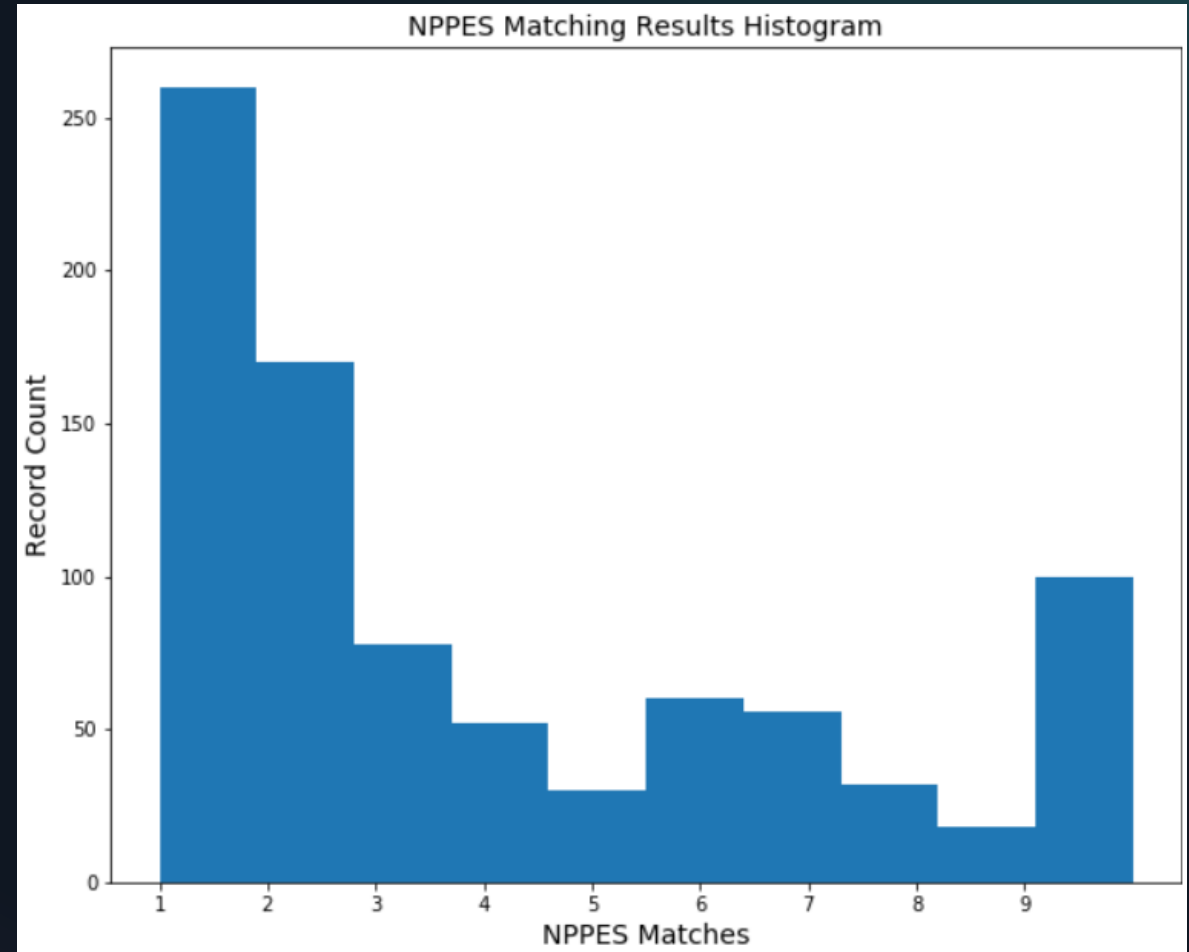
Step 4: Opioid Provider Dataset

- Final Aggregate Dataset: *704,463* unique providers based on NPI

- Year removed as dataset feature
  - Averages computed for all opioid-related features for timeframe
  - Difficult to determine exact timeframe of fraud based on NLP analysis
  - Avoids assumption provider acting fraudulently in all reporting years
  - Reduces complexity of analysis

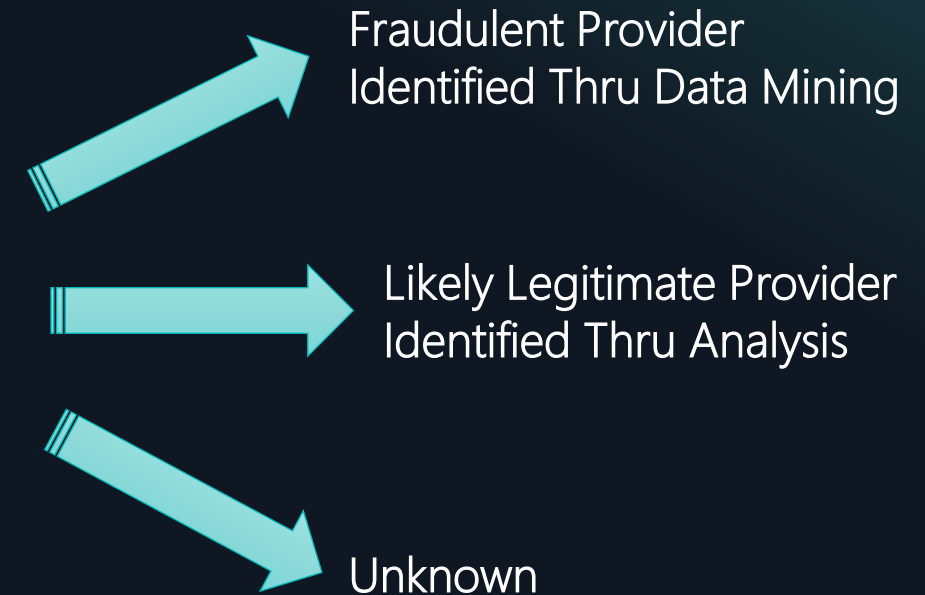| Feature	Mean
Total Claims	1895.568022
Total Beneficiaries	180.596029
Opioid Claims	124.341387
Opioid Beneficiaries	38.609345
Long Acting Opioid Claims	15.301676
Long Acting Opioid Beneficiaries	2.514243

Step 5: Likely Legitimate Opioid Provider

- Identification of Likely Legitimate Opioid Providers necessary for fully classified dataset

- Assumptions necessary to make determination not initially taken into consideration

- Mechanism to determine Likely Legitimate Opioid Providers will be established prior to execution of predictive models in Part 4

Predictive Model Preparation

- Establish Classifier Structure

    - Provider Classification

        - "Legitimate" (value = 0)
        - "Fraudulent" (value = 1)

    - Certainty Level

        - "Unknown" (value = 0)
        - "Known" (value = 1)

    - Two classifiers used to account for uncertainty / assumption of likely legitimate providers
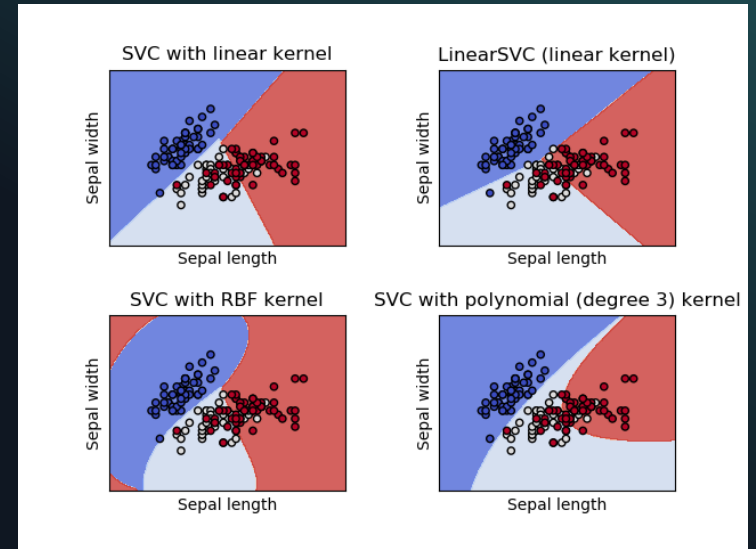
Predictive Model Preparation

Labeled Opioid Provider			
NPI	*{Other Characteristics}*	Provider Classification	Certainty Level
1111111111	...	1	1
2222222222	...	0	0
3333333333	...		

Fraudulent Provider Identified Thru Data Mining

Likely Legitimate Provider Identified Thru Analysis

Unknown

Predictive Model Preparation

- Key Initial Predictive Models
  - Linear Support Vector Classification
  - Naïve Bayes
  - Other Models will also be evaluated

- Modelling Platform
  - Orange
  - Auto-ML

Next Steps

Overview of Planned Part 4 Tasks:

- Implement Data Mining / NPPES Matching Improvements

- Address Likely Fraudulent Provider Identification

- Construct Predictive Model Workflow

- Execute Prediction & Interpret Results

- Compose Final Project Analysis

THANK YOU!

William Rubin

Email:
wrubin1@umbc.edu

GitHub:
https://github.com/warubin410/DATA606Spring2020

Google Site:
https://sites.google.com/umbc.edu/data606/home/willie-rubin