

**Predictive Modeling of Opioid Prescription Fraud
by Centers for Medicare and Medicaid Services (CMS) Providers**

Part 1: Problem Definition

Natural and synthetic opioids are powerful and highly effective prescription medicines used by doctors / health care providers in the treatment of moderate to severe pain in patients. Although useful as an approach to pain management, opioids have also shown to be dangerous as they can lead to addiction, overdoses, and even death. With a reported 46,000 opioid-related deaths in 2017 (an average of around 130 people per day) and an estimated 11.4 million people reportedly misusing prescriptions¹, the over-prescription of opioids presents a sizeable challenge to the healthcare system and its' patients in the United States.

As last publicly reported by the Centers for Disease Control (CDC) in 2017, the opioid prescription rate per 100 people in the United States was 58.7, which represents a significant decline from its peak of 81.3 in 2012². While experiencing a decline on the national level, on a local / community-level the popularity opioids has fostered an environment where providers and organizations can financially benefit from the over-prescription and misuse of the drugs for the treatment of pain.

The goal of this project is to evaluate the prescription patterns of providers which have been either formally charged, plead guilty to, or convicted by the United States Government of engaging in opioid-related fraud against CMS. By determining the key indicators exhibited by providers known to have committed fraud, this project aims to develop a predictive model to determine the likelihood of fraud being perpetrated by other providers.

Data Sources

The predictive model will be based upon analysis of the following sources of relevant data:

- [Department of Justice Press Releases](#) (2010-2020)
The Department of Justice issues regular press releases pertaining to individuals and organizations which have been formally charged, plead guilty to, or convicted of fraudulent activity. This project will focus on press releases identified as "Health Care Fraud" and reference opioids or opioid-related terms (such as specific drug names) for identification of provider / organization names.
- [National Provider Identifier \(NPI\) Registry](#)
The CMS National Plan and Provider Enumeration System (NPPES) contains a registry of all Medicare / Medicaid eligible providers. This project will utilize NPIs as a mechanism to accurately trace fraudulent provider / organization names to the records contained within the Medicare Provider Utilization and Payment Dataset.
- [Medicare Provider Utilization and Payment Data: Part D Prescriber Summary Tables](#) (2013-2017)

The Public Use Files (PUF) contain information on prescription drugs prescribed by providers which have been paid for using the Medicare Part D Prescription Drug Program. The data is based upon the CMS Chronic Conditions Data Warehouse with provider-based prescription records associated with Medicare Advantage Prescription Drug (MAPD) plans and Prescription Drug Plans (PDP). Records are aggregated on a per year basis by unique National Provider Identifiers (NPI) and include the following categories of data: total number of prescriptions dispensed, total drug cost, beneficiary characteristics, and drug categories. This data will be utilized in my project to determine provided-based prescription trends.

- [Physician Compare 2017 Individual Eligible Clinician Public Reporting - Overall MIPS Performance](#)

The Physician Compare file contains the Merit-Based Incentive Payment System (MIPS) final scores and performance category scores for all participating clinicians in 2017. This data will be utilized in my project to determine likely legitimate opioid providers.

Part 2: Existing Research and Initial Exploratory Data Analysis

Existing Research Studies

Two existing research studies were identified which are directly relevant to my projects' main goal of developing a predictive model to determine the likelihood of opioid-related fraud being perpetrated by a Medicare Part D Provider. First, the US Department of Health and Human Services Office of Inspector General (HHS-OIG) conducted a study that evaluated the prescription drug events records from 2016 with the goal of protecting beneficiaries from the adverse effect of opioid abuse³. The study's three main components were: an analysis of opioid dispensation and spending (utilization) across all beneficiaries; individual beneficiary usage accounting for prescription strength and treatment duration; and identification of providers which exhibited extreme prescription patterns.

Key results of this study include that one-third of Medicare Part D beneficiaries received at least one opioid prescription; approximately 1/2 million beneficiaries received large opioid amounts; and around 400 providers exhibited potentially questionable prescription practices.

The second identified research effort is a CMS developed interactive mapping tool which enables users to perform comparisons of Medicare Part D opioid prescriptions on a national-level scale⁴. The tool allows users to explore claim totals and prescribing rates for both opioid and long-acting opioid prescriptions from 2013 through 2017. Percentage change is also determined and displayed for the entire timeframe to assist in visualization of trends at both the state and county levels.

Although in alignment with the referenced studies, the goal of this project would expand upon scope to include an analysis based on known cases of opioid prescription related fraud. By

including this aspect, this project endeavors to establish a predictive capability to actively identify suspect providers.

Initial Exploratory Data Analysis

The Medicare Provider Utilization and Payment Data: Part D Prescriber Summary Tables⁵ for 2013 through 2017 were downloaded and analyzed utilizing Python within JupyterLabs. The resulting combined data frame contained 84 source columns / features⁶ with an additional column added during load for calendar year, which is identified in the filename. There exists one record for each provider with a registered NPI for each year within the timeframe publicly available, which results in approximately 5.5 million source records.

Each record contains four categories of data pertaining to the provider within that specified calendar year. These categories include details on the following: provider registration, applicable Medicare Program, prescription drug utilization, and beneficiary demographics / health characteristics.

Some key outcomes of initial exploratory data analysis are:

- Overall Provider Characteristics (Not Opioid Specific)
 - Top 10 States by Number of Providers
 - California has largest number of Medicare Part D Providers (143,640), which is not unexpected due to its large population.
 - Florida ranks #4 (82,214) despite an overall large number of Medicare Recipients and is significantly less than the top 2 (New York and California).
 - Top 10 Provider Specialties
 - Nurse Practitioner was the highest ranked specialty (178,337).
 - Overall top 6 ranked specialties (Nurse Practitioner, Dentist, Internal Medicine, Student in an Organized Health Care Education / Training Program, Family Practice and Physician Assistant) seem reasonable but students ranking of #4 (118,911) is an interesting finding.
 - Drop-off between Physician Assistant (#6 - 104,851) and Emergency Medicine (#7 - 48,943) is quite drastic and surprising.
- Opioid Specific Provider Characteristics
 - Opioid Beneficiary Ratio by State (Number of Opioid Beneficiaries to Total Beneficiaries)
 - Top 5 ranking of Alabama, Arkansas, Tennessee, Mississippi, Oklahoma all coincide with the 2017 HHS-OIG Study, although the order varies slightly.
 - New York and Hawaii ranked as two of the bottom three also coincides with the 2017 HHS-OIG Study (Puerto Rico was included in my analysis but may have been excluded from the HHS-OIG Study).
 - Opioid Beneficiary Ratio by Provider Specialty
 - Although Specialist/Technologist Cardiovascular has highest ratio (0.916667), the calculation is based on a very small beneficiary total (12).

- Interventional Pain Management (0.786231) and Pain Management (0.768426) seem reasonable rankings as the top 2 with large beneficiary base (2,524,681 and 2,301,019 respectively).
- Out of the top 25 identified specialties which had opioid beneficiaries, the difference between Thoracic Surgery (#25 - 0.419788) and Interventional Pain Management (#2 - 0.786231) appear quite drastic (0.366443).
- High-level Opioid-Specific Claims / Beneficiary Analysis
 - National Part D Beneficiary Totals / Trends
 - Total beneficiaries experienced a fairly steady climb throughout the timeframe.
 - Beneficiaries receiving opioids rose until a peak in 2015 and then declined to a similar number occurring in 2013.
 - Beneficiaries receiving long-active opioids was fairly flat until 2015 and then declined at a steady rate.
 - Rate at which total beneficiaries change slightly from 2013 to 2017 but continually increased.
 - Rate at which beneficiaries received opioids declined steadily from 2013 to 2017.
 - Rate at which beneficiaries received long-acting opioids slowed from 2013 to 2015 and then drastically declined from 2015 to 2017.
 - National Part D Claim Totals / Trends
 - Total claims experienced a fairly steady climb throughout the timeframe (similar to beneficiary totals).
 - Claims including opioids increased until 2014 and then slowly declined (peak occurred one year earlier than beneficiary totals).
 - Claims including long-active opioids was fairly flat until 2015 and then declined (similar to beneficiary totals).
 - Rate at which total claims increased followed a similar pattern as beneficiary totals but at lower percentages.
 - Rate at which opioid-related claims changed did not follow a similar pattern as the beneficiary totals, with a dip 2015, followed by an increase in 2016 and then a sharp decline.
 - Rate at which long-acting opioid-related claims changed followed a similar pattern as beneficiary totals but at slightly lower rates.
- Opioid and Long-acting Opioid Prescriber Rate Histogram
 - Provider prescriber rate for opioids demonstrates highest totals around 5% and then drops off sharply until around 15% before a steady decline.
 - Provider prescriber rate for long-acting opioids demonstrates highest totals around 10% and then slowly drops off until around 50% before a steady decline.
- Opioid and Long-acting Opioid Prescriber Cost Analysis
 - $(\text{Average Opioid Cost} = \text{Total Opioid Drug Cost} / \text{Total Opioid Claims})$ on a provider basis
 - Of the top 50 providers with highest average opioid cost (range of 1380.62 - 5778.87) the most frequent specialties were Internal Medicine (20%), Nurse Practitioner (16%), Hematology/Oncology (16%), and Radiation Oncology (12%).

- (Average Long-Acting Opioid Cost = Total Long-Acting Opioid Drug Cost / Total Long-Acting Opioid Claims) on a provider basis
- Of the top 50 providers with highest average opioid cost (range of 2239.12 - 7610.47) the most frequent specialties were Internal Medicine (30%), Family Practice (18%), Nurse Practitioner (14%), Physical Medicine and Rehabilitation (6%), and Hematology/Oncology (6%).
- Family Practice exhibited a large increase from 8% for opioid to 18% for long-acting opioids.

Part 3: Final Exploratory Data Analysis and Predictive Model Preparation

The main goal of this part was to prepare the data for use with machine learning algorithms and predict Medicare Part D Providers which may be fraudulently prescribing opioids. In order to accomplish this objective, a labeled dataset was constructed which contained all CMS Providers identified in the Part D Prescriber Summary Table as having provided opioids to patients. The opioid provider dataset was then augmented to include a classifier to indicate fraudulent / legitimate practices in addition to a indication of level of certainty.

The following tasks were completed to create the CMS Opioid Provider Labeled Dataset (depicted in Figure 1 below), which will be utilized in Part 4 to predict the likelihood of prescription related fraud:

- Developed Data Pipeline for Fraudulent Provider Identification
 - Data Collection: DOJ Press Release website, filtered for opioid-related cases, was scraped utilizing Scrapy to compile a corpus of text associated with CMS providers formally charged, plead guilty to, or convicted of fraudulent activity. When executed on 2020-03-11, the data collection Python code retrieved **120 press releases** from the DOJ website.
 - Entity / Location Extraction: Dataset of each scraped opioid-related DOJ Press Release text was cleansed and spaCy (NLP Python Library) was utilized to extract entities from the text. The spaCy library supports extraction of numerous entities, for my purposes the entities extracted included: ORG (organization), PERSON (names of people), GPE (country, city, state), and LOC (non-GPE location).
 - NPI Resolution: For all people (first and last name) and states extracted from each scraped press release, API calls were constructed to lookup and retrieve any officially assigned identifiers from the [NPPES NPI Registry](#) matching the parameters. Of the **overall 856** matches found, **260 resulted in exact matches** and were then included in the initial known fraudulent provider dataset. For Part 4 of this project, this process will be improved to handle instances when multiple potential matches are returned from the NPPES NPI Registry API Lookup.
- Performed Final EDA of Part D Prescriber Summary Table Data

- Opioid Provider Dataset: Created aggregated dataset of **704,463 providers** based on NPI which reported values for opioid or long-acting opioid characteristics. Opioid provider analysis conducted was based upon records grouped by NPI across the five years contained within the Summary Table Data. This decision was made due to time constraints of extracting relevant timeframes in which the fraud occurred from DOJ Press Releases using NLP.
- Likely Legitimate Provider Determination: Analyzed opioid provider dataset to identify averages and outliers in terms of opioid and long-acting opioid characteristics (Beneficiary Count, Claim Count, Prescription Rate, Drug Cost, and Day Supply). Of note, the assumptions necessary to determine likely legitimate providers was not initially taken into consideration but is necessary to properly classify providers in the labeled dataset. This determination will be revised prior to execution of predictive models in Part 4.
- Predictive Modeling Preparation
 - Establish Classifiers: Determined the following two classifiers to be used in the creation of the labeled dataset: “Provider Classification” will be used to distinguish between “Legitimate” (value = 0) and “Fraudulent” (value = 1); “Certainty Level” will be used to identify “Unknown” (value = 0) and “Known” (value = 1). Based on these classifiers, records with values of (1,1) would indicate fraudulent providers identified through the data mining process and those with values of (0,0) would indicate providers determined to be likely legitimate.
 - CMS Part D Opioid Provider Labeled Dataset Creation: Combined Opioid Provider Dataset with classified listing of fraudulent and likely legitimate providers determined in prior steps to compile labeled data. Classifiers for remaining providers will be left blank to indicate “unknown” status.
 - Model Construction: Based upon initial research and evaluation of the [scikit-learn algorithm cheat sheet](#), focus will be placed on implementation of Linear Support Vector Classification (SVC) and Naive Bayes. The analysis will be performed using [Orange](#) for rank and feature selection, split of data between testing and training, and algorithm execution. Other algorithms will also be explored and evaluated in terms of applicability and accuracy to produce the most statistically significant results.

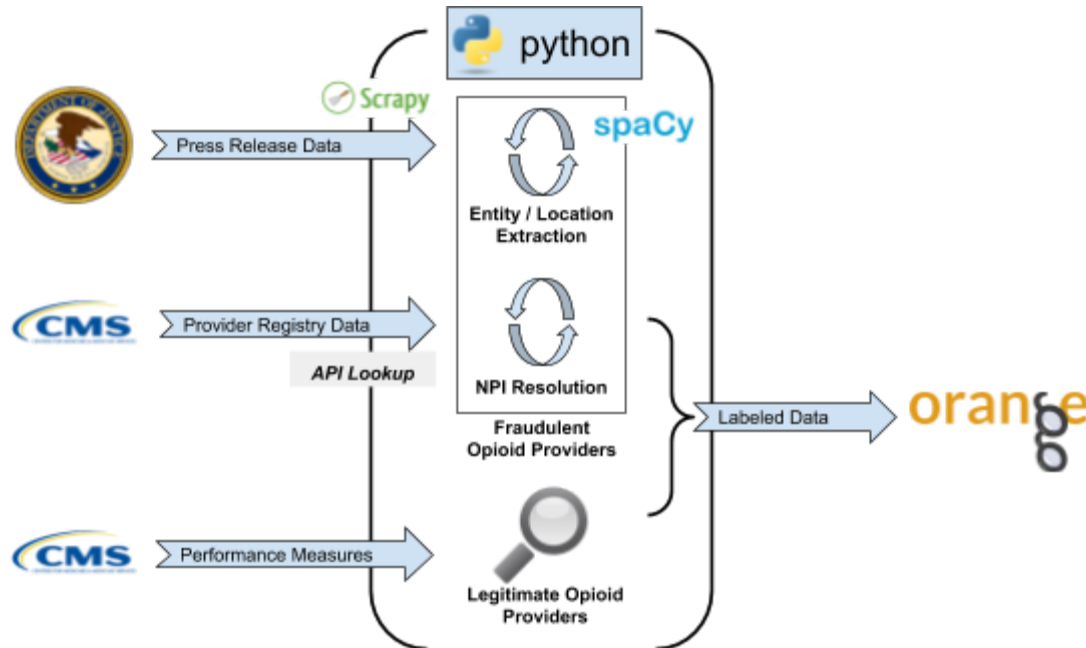


Figure 1. CMS Part D Opioid Provider Labeled Dataset Process

The following table provides details pertaining to the additional libraries and products utilized in this project:

Table 1. CMS Part D Opioid Provider Labeled Dataset Technology Stack

Product / Library Name	URL	Purpose
Scrapy	https://scrapy.org	Open-source Python library which provides a framework to facilitate the extraction of press release text data from the DOJ Website.
spaCy	https://spacy.io/	Open-souce Python library which provides Natural Language Processing (NLP) capabilities and utilized to perform entity extraction (names and locations) from corpus of DOJ Press Release scraped text.
Orange	https://orange.biolab.si/	Open source Python based environment providing robust machine learning and data visualization capabilities for executing predictive analysis.

Part 4: Execution and Interpretation

Final Labeled Dataset Adjustments

Before executing the prediction modeling and interpretation of results, several adjustments were needed to finalize the labeled dataset of opioid providers. First, the Fraudulent Provider Identification Data Pipeline was modified to address instances where multiple NPIs were identified during the resolution process. Unfortunately while completing this effort, it was determined several of the 120 DOJ Press Releases on which data mining was originally performed were not fraud-related. The press releases in question were policy-based and thus resulted in an inflated number of fraudulent NPIs identified in Part 3.

To address this issue, the data pipeline was modified to include extraction of identified verbs within the text. The additional extracted data was used to filter out press releases which were unlikely fraud related (see Appendix A for verb listing). This resulted in removal of a subset of extracted entities associated with nine press releases originally included in Part 3.

To resolve the instances of multiple NPIs being returned based on the remaining identified names, the extracted location processing was enhanced to include city-like names (see Appendix A for city filtering terms). The extracted city names were then included in the NPPES API call to narrow results. After implementation of these adjustments, removal of duplicates, and final manual verification of inclusion, the following table outlines the achieved results of the fraudulent provider data pipeline efforts:

Table 2. Fraudulent Provider Data Pipeline Results

DOJ Press Releases	111
Extracted Entities / Terms	21,113
Resolved NPIs	221

The second labeled dataset adjustment needed was identification of likely legitimate CMS Part D Opioid Providers. To determine a subset of CMS Part D Providers which were likely to be acting legitimately in prescription practices, a new dataset pertaining to the CMS Quality Payment Program (QPP) was identified and incorporated. The dataset pertains to the Merit-based Incentive Payment System (MIPS), which is one component of the QPP that rewards participant clinicians with payment adjustments based on their cost efficiency, quality of care, and health outcomes⁷.

Since participation in MIPS is voluntary, adds additional scrutiny to a providers practices, and can either positively or negatively affect payments from CMS, it is highly unlikely a fraudulent provider would participate. Thus, for the purposes of identifying likely legitimate opioid providers

for this project, the published MIPS Final Scores for 2017 where providers received a perfect score (100) were merged with the opioid provider dataset. No instances of providers identified as fraudulent were identified in the filtered MIPS-related data.

The table below details the final composition of the CMS Part D Opioid Provider Labeled Dataset utilized in predictive modelling efforts.

Table 3. CMS Part D Opioid Provider Labeled Dataset Composition

All Opioid Providers	704,463
Identified Fraudulent Opioid Providers	85* * 221 identified overall but only 85 remained after cross-reference to opioid provider dataset, likely due to inactive status
Likely Legitimate Opioid Providers	38,229

Predictive Modeling

Based upon the composition of the final CMS Part D Opioid Provider Labeled Dataset, a semi-supervised machine learning approach was determined to be the best option. Semi-supervised learning falls between supervised (with labeled training data) and unsupervised (with no labeled training data) approaches and provides a practical method to improve learning accuracy when only a small amount of labeled data is available⁸.

To utilize the semi-supervised learning approach for the project, pseudo-labeling was employed to assign approximate labels to the unlabeled portion of the dataset⁹. First, the labeled portion of the dataset was separated and used to train an initial model. The resulting initial model was then used against the remaining dataset to predict fraudulent labels for the previously unlabeled data. Finally, the two datasets were merged and used to create the final predictive model.

Results

The semi-supervised learning approach was implemented within the Orange3 Data Analysis Environment (see Appendix B to view entire prediction model workflow), the following results were observed for each step of the process:

- **Labeled Provider Modeling:** After filtering the opioid provider dataset for the 38,314 labeled providers, the Information Gain Ratio algorithm was used to determine the top 10 impactful features. These features included (in descending order of importance):
la_opioid_bene_count, la_opioid_drug_cost,
average_age_of_beneficiaries, opioid_prescriber_rate,
la_opioid_claim_count, beneficiary_race_nat_ind_count,
antipsych_bene_count_ge65, benecificiary_age_less_65_count,
opioid_claim_count.

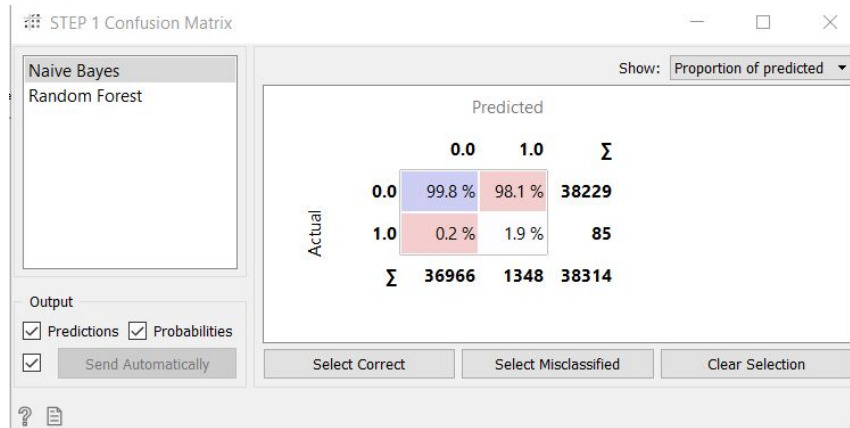


Figure 2. Labeled Opioid Provider Confusion Matrix: Naive Bayes

- Pseudo-Labeling of Unlabeled Providers: The Naive Bayes Model generated above was then applied to the remaining 666,149 unlabeled providers. This pseudo-labeling process resulted in creation of approximate / predicted labels for the previously unknown classifiers.
- Combined Dataset: The labeled and pseudo-labeled datasets were then concatenated to produce the final fully labeled opioid provider dataset.

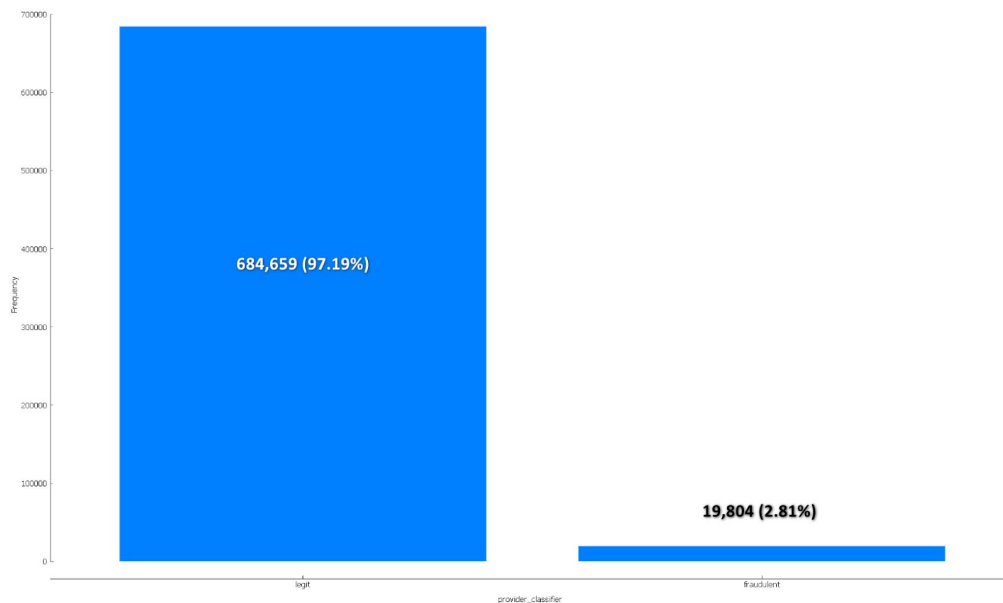


Figure 3. Fully Labeled Opioid Provider Distribution

- Fraudulent Provider Prediction Modeling: The fully labeled opioid provider dataset was then sampled using two approaches with the following results:

- 70/30 Split - The table below display the outcomes of the predictive models based on the 70/30 sampling split. Although Neural Network did perform slightly better, the low number of fraudulent labeled providers lead to the identification of no instances of a fraudulent provider. Thus, only the confusion matrix for Naive Bayes was included, and depicted in the figure below.

Table 4. CMS Part D Opioid Provider Labeled Dataset Composition

Model Type	AUC	F1	Precision
Naive Bayes	0.934	0.918	0.191
Neural Network	0.958	0.972	0.121

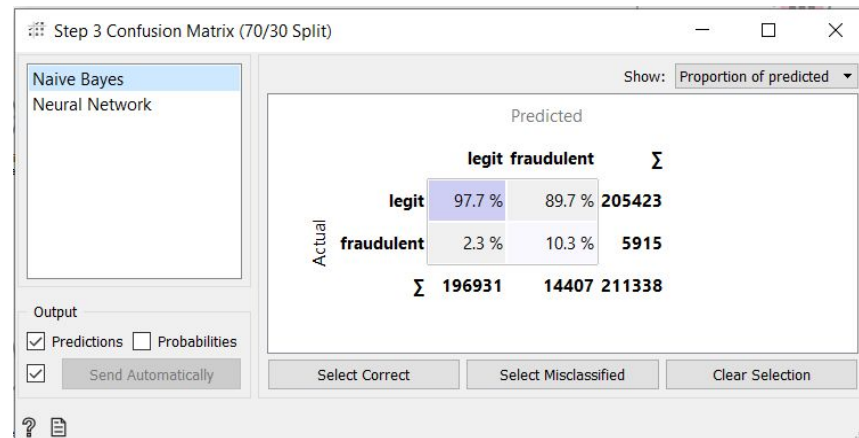


Figure 4. Prediction Model (70/30 Split) Confusion Matrix: Naive Bayes

- Cross Validation (10 Folds) - The table below display the outcomes of the predictive models based on cross validation (10 folds) sampling. Neural Network performed the same and Naive Bayes improved slightly.

Table 5. CMS Part D Opioid Provider Labeled Dataset Composition

Model Type	AUC	F1	Precision
Naive Bayes	0.944	0.935	0.165
Neural Network	0.958	0.972	0.121

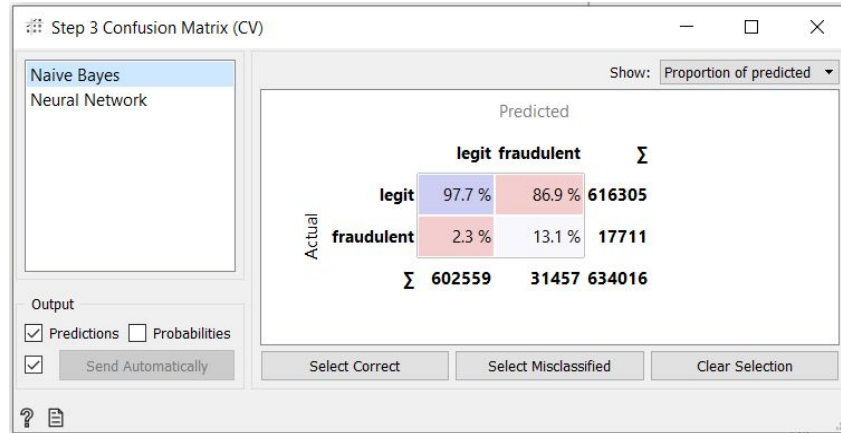


Figure 5. Prediction Model (CV) Confusion Matrix: Naive Bayes

Conclusion

Although the semi-supervised learning approach with pseudo-labeling yielded results the model prediction was overall hampered by the small size of identified fraudulent providers. I believe with additional data sources and enhancements to the NLP analysis, the results of the fraudulent data pipeline would improve. In conjunction with the process implemented to create the CMS Part D Opioid Provider Labeled Dataset Process, this would position the developed prediction workflow to produce more actionable results.

Lessons Learned

- Early data mining assumption that simple extraction of names and locations from the filtered set of well-structured DOJ Press Releases would not yield a significant amount of falsely identified fraudulent providers was inaccurate. Although lawyers, prosecutors, etc. did not have registered NPIs, the commonality of extracted names amongst all identified locations lead to a significant number of misidentifications since position in text was not considered. Also, a number of press releases were policy or informational based which added to misidentification.
- Basic utilization of the spaCy NLP Python Library to extract entities, locations, and parts of speech was powerful and relatively straightforward. Contextual analysis of the press releases and identification of persons of interest and associated locations was quite challenging. This resulted in use of only rudimentary NLP analysis and contributed to false identification of fraudulent providers.
- All necessary components of the opioid provider labeled dataset were not considered at project on-set. Identification of likely legitimate opioid providers was initially overlooked and lead to delays in the predictive modeling preparation steps.

Future Considerations

- Identify additional data sources for use in identification of fraudulent opioid providers to increase the number of labeled instances and improve prediction model / results.

- Improve NLP to consider an entities position within the text to reduce misidentification and incorporate better contextual processing of text to avoid data mining of informational-only and policy-based press releases.
- Improve determination of likely legitimate opioid providers beyond use of MIPS Final Scoring to include additional datasets.
- Incorporate time component into analysis to evaluate opioid providers trends throughout the span of publicly available data (2013-2017).
- Extract date and timeframes from DOJ Press Releases to better map provider fraud to specific years and provider more granular analysis.
- Improve Fraudulent Providers Data Mining Pipeline to include determination of providers that have been charged multiple times and develop a weighting mechanism so they will be more strongly classified.
- Incorporate feature engineering to improve prediction results.

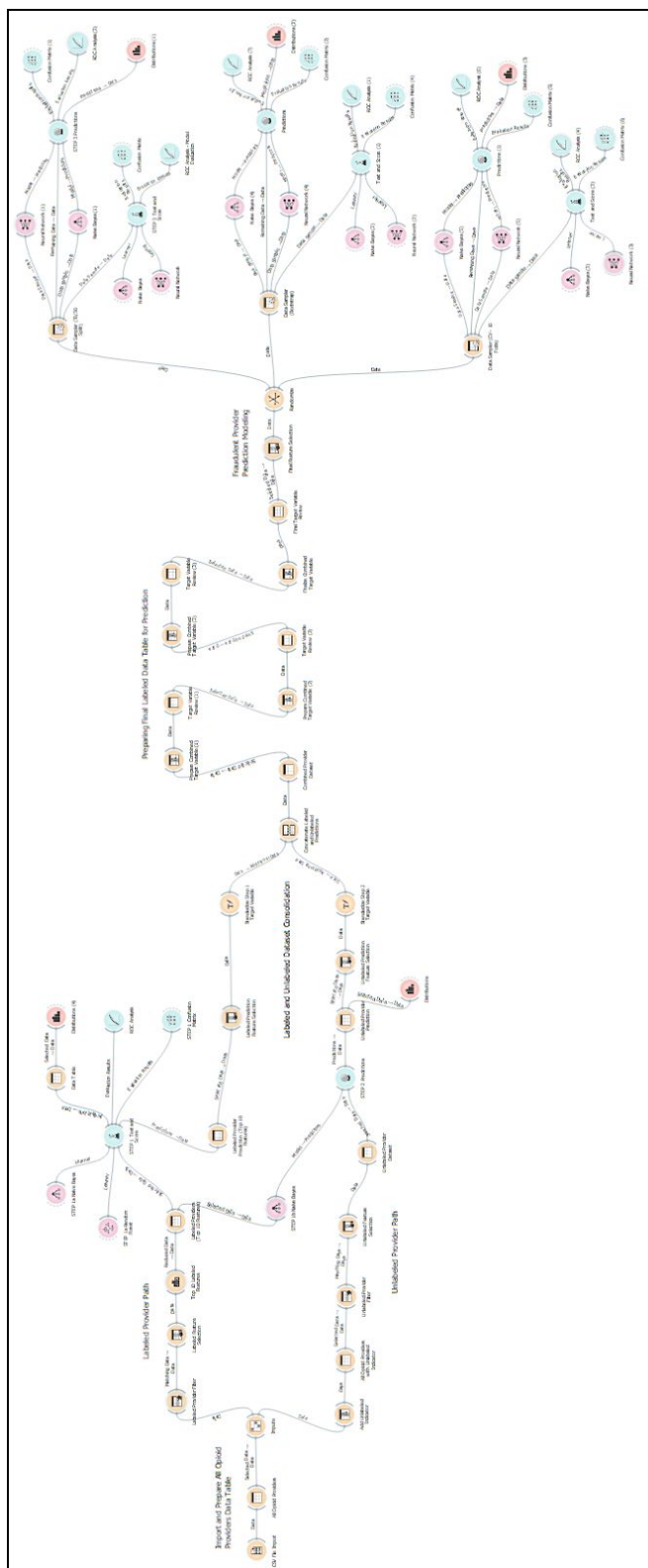
References

- ¹ CMS, "CMS Roadmap: Strategy to Fight the Opioid Crisis." April 2020. Retrieved from <https://www.cms.gov/About-CMS/Agency-Information/Emergency/Downloads/Opioid-epidemic-roadmap.pdf> on 2020-02-12.
- ² CDC, "U.S. Opioid Prescribing Rate Maps." March 2020. Retrieved from <https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html> on 2020-02-12.
- ³ HHS-OIG, "Opioids in Medicare Part D: Concerns about Extreme Use and Questionable Prescribing", OEI-02-17-00250, July 2017. Retrieved from <https://oig.hhs.gov/oei/reports/oei-02-17-00250.asp> on 2020-02-24.
- ⁴ CMS, "Medicare Part D Opioid Prescribing Mapping Tool / Methodology", April 2019. Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Downloads/Opioid_Methodology.pdf on 2020-02-24.
- ⁵ CMS, "Medicare Provider Utilization and Payment Data: Part D Prescriber Summary Tables" available at <https://data.cms.gov/browse?category=Medicare+-+Part+D&sortBy=alpha&tags=provider+summary&utf8=%E2%9C%93>
- ⁶ CMS, "Medicare Fee-For Service Provider Utilization & Payment Data Part D Prescriber Public Use File: A Methodological Overview", April 2019. Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Downloads/Prescriber_Methods.pdf on 2020-02-20.
- ⁷ CMS, "Quality Payment Program: MIPS Overview." Retrieved from <https://qpp.cms.gov/mips/overview> on 2020-04-25.
- ⁸ Wikipedia contributors. "Semi-supervised learning." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 2020-04-19. Retrieved from https://en.wikipedia.org/wiki/Semi-supervised_learning on 2020-05-09.
- ⁹ Jain, Shubham. "Introduction to Pseudo-Labeling : A Semi-Supervised learning technique." *Analytics Vidhya*. 2017-09-21. Retrieved from <https://www.analyticsvidhya.com/blog/2017/09/pseudo-labelling-semi-supervised-learning-technique/> on 2020-05-09.

Appendix A: NLP Entity Extraction Filtering

Entity Type	Key Terms	Usage
Verb	pled, pleaded, conspire, conspired, arraigned, alleged, charged, distribute, distributed, investigated, investigate, prosecute, prosecuted	Include Terms
Location (GPE)	"U.S.", "the United States", "M.D.", "US", "America", <state names>, <state codes>	Exclude Terms

Appendix B: Orange3 Prediction Model Workflow



Appendix C: Project-Related Materials

Purpose	Link
Source Code and Project Artifact Repository	https://github.com/warubin410/DATA606Spring2020
Video Presentations	<p>Delivery 1: https://youtu.be/aRNdMIUhwXI</p> <p>Delivery 2: https://youtu.be/PPh2ywGZHEM</p> <p>Delivery 3: https://youtu.be/okvpX_x4sSY</p> <p>Delivery 4: https://youtu.be/bZBCfmrnX6E</p>
Project Site / Overview	https://sites.google.com/umbc.edu/data606/home/willie-rubin