Predictive Analysis in Identifying Frequent Opioid Prescriber

Erjia Meng, Jiayi Hao, Shiyu Liu, Xinyi Liu, Yulin Feng

Opioid & Opioid Epidemic

- Prescription Opioids - a class of drugs used treat moderate to severe pain
- Opioids are highly addictive, and also have serious risks and side effects
- Opioids were involved in 49,860 overdose deaths in 2019

  (70.6% of all drug overdose deaths).

https://www.hopkinsmedicine.org/opioids/what-are-opioids.html

Opioid Over-Prescribers & “Pill Whales”

Pill Whale - Doctors who are over-prescribing Opioid drugs, and thus gain profit from medical representative.

First Step to find over-prescriber:

Identify Frequent Opioids Prescriber

Main Goal of our project

https://ethicsunwrapped.utexas.edu/video/whale-watching

Identify Frequent Opioids Prescriber

Predictive Model:

- Systematic method with Machine Learning techniques
- Dataset: 25,000 unique licensed medical professionals in the United States
- Independent Variable - Basic information of the medical professionals
- Target Variable - whether the professional is a 'Frequent Opioid Prescriber'

Frequent Opioid Prescriber ≠ Over-Prescriber

→ Allow further investigation to identify over-prescribers

Datasets

Datasets

- Main dataset prescriber-info.csv : basic information of some medical professionals in 2014 in the United States and their prescription records for hundreds of common opioid and non-opioid drugs
- overdoses.csv : contains information on opioid related drug overdose fatalities

Source: https://www.kaggle.com/datasets/apryor6/us-opiate-prescriptions

Datasets

prescriber-info.csv

- NPI - unique National Provider Identifier number
- Gender - M/F
- State - U.S. State by abbreviation
- Credentials - set of initials indicative of medical degree
- Specialty - description of type of medicinal practice
- A long list of drugs - with numeric values indicating the total number of prescriptions written for the year by that individual
- Opioid.Prescriber - a boolean label indicating whether or not that individual prescribed opiate drugs more than 10 times in the year

Datasets

overdose.csv

- State - full name of U.S. State
- Population - population every state
- Death - number of death caused by overdose
- Abbrev - U.S. State by abbreviation

Clean data

prescriber-info

- NPI - unique National Provider Identifier number ← drop
- Gender - M/F ← To digit
- State - U.S. State by abbreviation ←
- Credentials - set of initials indicative of medical degree
- Specialty - description of type of medicinal practice
- A long list of drugs with numeric values indicating the total number of prescriptions written for the year by that individual
- Opioid.Prescriber - a boolean label indicating whether or not that individual prescribed opiate drugs more than 10 times in the year

Overdose.csv

- State - full name of U.S. State
- Population - population every state
- Death - number of death caused by overdose
- Abbrev - U.S. State by abbreviation

Death per capita

Dealing with “State” feature

State feature from prescriber_info.csv

Washington D.C. 79 rows

Puerto Rico 231 rows

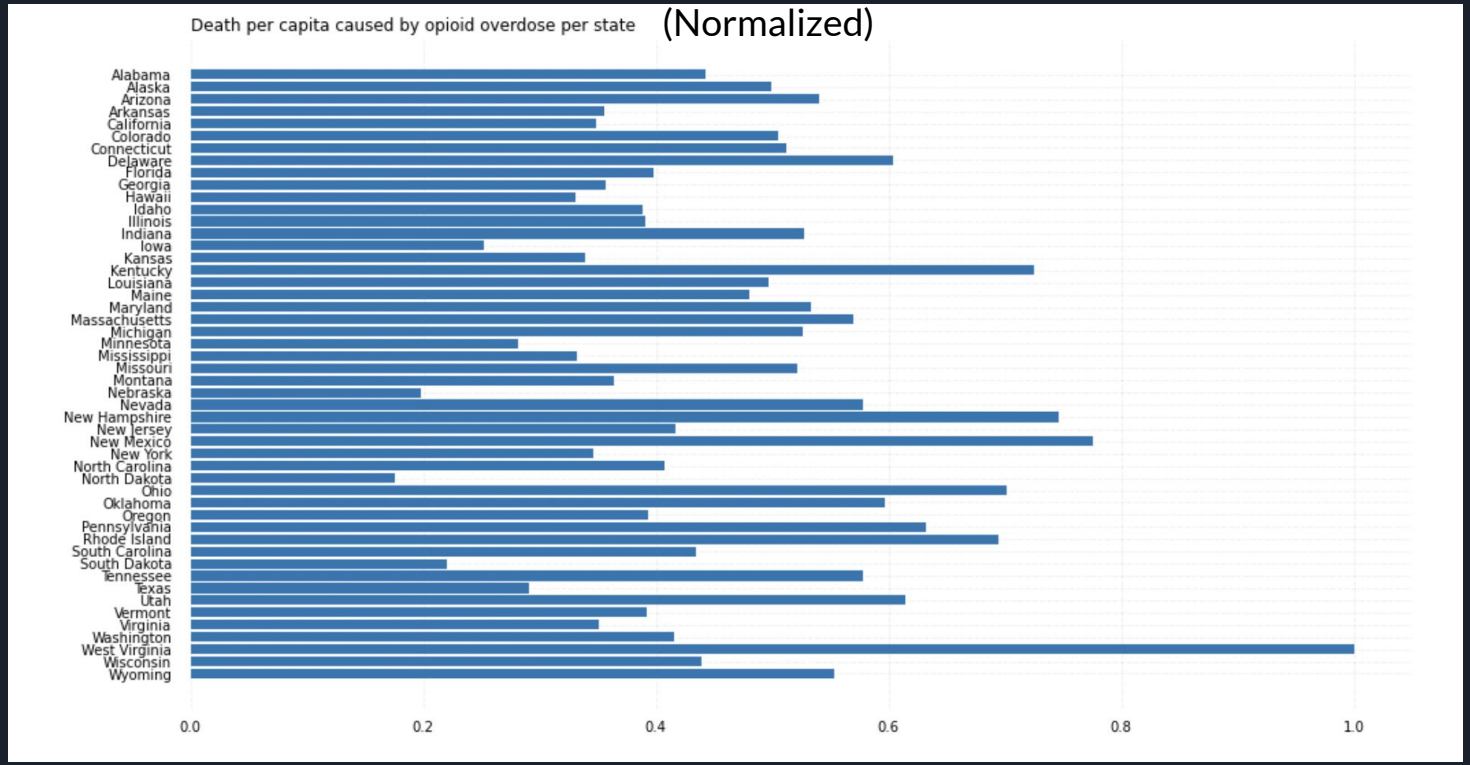
```
np.sort(arr1)
```

```
array(['AA', 'AE', 'AK', 'AL', 'AR', 'AZ', 'CA', 'CO', 'CT', 'DC', 'DE',
       'FL', 'GA', 'GU', 'HI', 'IA', 'ID', 'IL', 'IN', 'KS', 'KY', 'LA',
       'MA', 'MD', 'ME', 'MI', 'MN', 'MO', 'MS', 'MT', 'NC', 'ND', 'NE',
       'NH', 'NJ', 'NM', 'NV', 'NY', 'OH', 'OK', 'OR', 'PA', 'PR', 'RI',
       'SC', 'SD', 'TN', 'TX', 'UT', 'VA', 'VI', 'VT', 'WA', 'WI', 'WV',
       'WY', 'ZZ', dtype=object)
```

State feature from overdose.csv

```
np.sort(arr2)
```

```
array(['AK', 'AL', 'AR', 'AZ', 'CA', 'CO', 'CT', 'DE', 'FL', 'GA', 'HI',
       'IA', 'ID', 'IL', 'IN', 'KS', 'KY', 'LA', 'MA', 'MD', 'ME', 'MI',
       'MN', 'MO', 'MS', 'MT', 'NC', 'ND', 'NE', 'NH', 'NJ', 'NM', 'NV',
       'NY', 'OH', 'OK', 'OR', 'PA', 'RI', 'SC', 'SD', 'TN', 'TX', 'UT',
       'VA', 'VT', 'WA', 'WI', 'WV', 'WY'], dtype=object)
```


Death per capita caused by opioid overdose per state
(Normalized)
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
0.0
0.2
0.4
0.6
0.8
1.0

Features

Gender (0/1)

State (str)

Credentials (str)

Specialty (str)

List of Drugs (0/1)

Add a death_Rate_norm column

Death_Rate_norm for “DC” and “PR” is filled by mean value

Features

Gender (0/1)

State (str) ← **Categorical feature**

Credentials (str)

Specialty (str)

List of Drugs (0/1)

Death_Rate_norm (float between 0 and 1)

One-hot encoding on State

```
features_df.head()
```

index	Gender_is__M	State_is__AL	State_is__AR	State_is__AZ	State_is__CA	State_is__CO	State_is
0	1	0	0	0	0	0	
1	0	1	0	0	0	0	
2	0	0	0	0	0	0	
3	1	0	0	1	0	0	
4	1	0	0	0	0	0	

5 rows × 346 columns

Features

Gender (0/1)

List of State_is_ (0/1)

Credentials (str)

Specialty (str)

List of Drugs (0/1)

Death_Rate_norm (float between 0 and 1)

prescriber-info

- NPI - unique National Provider Identifier number
- Gender - M/F
- State - U.S. State by abbreviation
- Credentials - set of initials indicative of medical degree
- Specialty - description of type of medicinal practice
- A long list of drugs with numeric values indicating the total number of prescriptions written for the year by that individual
- Opioid.Prescriber - a boolean label indicating whether or not that individual prescribed opiate drugs more than 10 times in the year

Credentials

```
df.Clean_Credentials.unique()
```

```
array(['DDS', 'DMD', 'MD', 'NoRecord', 'DDS MS', 'DDSPLLC', 'DMD MS',
       'DDSMS', 'DDS PA', 'DDS MPH', 'DDSMAGD PC', 'DMDPC', 'DENTIST',
       'DD S', 'DDS PC', 'DDS PHD', 'DR', 'DMDFAGD', 'DDS MD',
       'DDS M                                                  ISTDDS',
       'DDS M                                                   MAGD',
       'HERSC                                                   'DMD MMSC',
       'BDS I                                                  DDS MS PHD',
       'DDS RN', 'APC DDS', 'BS DDS', 'DDS FACD', 'DMD MSD', 'EIN',
       'DDSPRACTICELIMITE', 'DDS GREGSAWYER', 'DMDMS', 'DMD PC',
       'DDS PHARMD', 'BDS', 'DDSPA', 'DDS PD', 'DO', 'DDSFAGD', 'DDSPS',
       'DMD MDS', 'DMDDRMEDDENT', 'DDS MS MSPH', 'MMD', 'DDS PS',
       'DDS PLLC', 'DDA', 'DMD MPH', 'DMDBA', 'DMDPA', 'MSDMD', 'OD',
       'ODPC', 'MD OD', 'DOCTOROFOPTOMETRY', 'OPTOMETRIST', 'DO DOS',
       'OO', 'FAAO OD', 'MED OD', 'FAAO', 'MS OD', 'FAAO FOAA OD',
       'MPH OD', 'OD PA', 'ODPS', 'MDPA', 'MDPC', 'MD MPH', ' MD',
```

```
len(df.Clean                                    ue())
```

```
618
```

Drop It

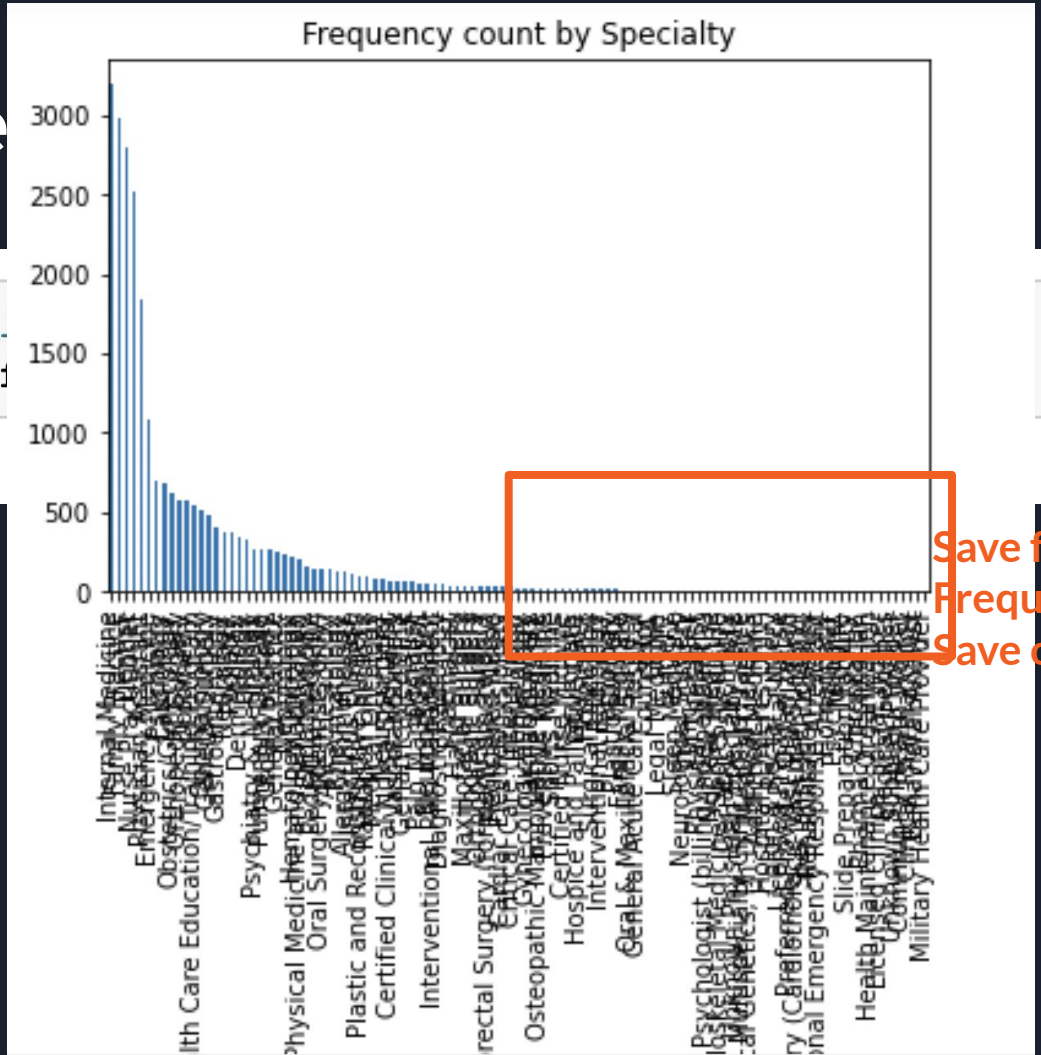
prescriber-info

- NPI - unique National Provider Identifier number
- Gender - M/F
- State - U.S. State by abbreviation
- Credentials - set of initials indicative of medical degree
- Specialty - description of type of medicinal practice
- A long list of drugs with numeric values indicating the total number of prescriptions written for the year by that individual
- Opioid.Prescriber - a boolean label indicating whether or not that individual prescribed opiate drugs more than 10 times in the year

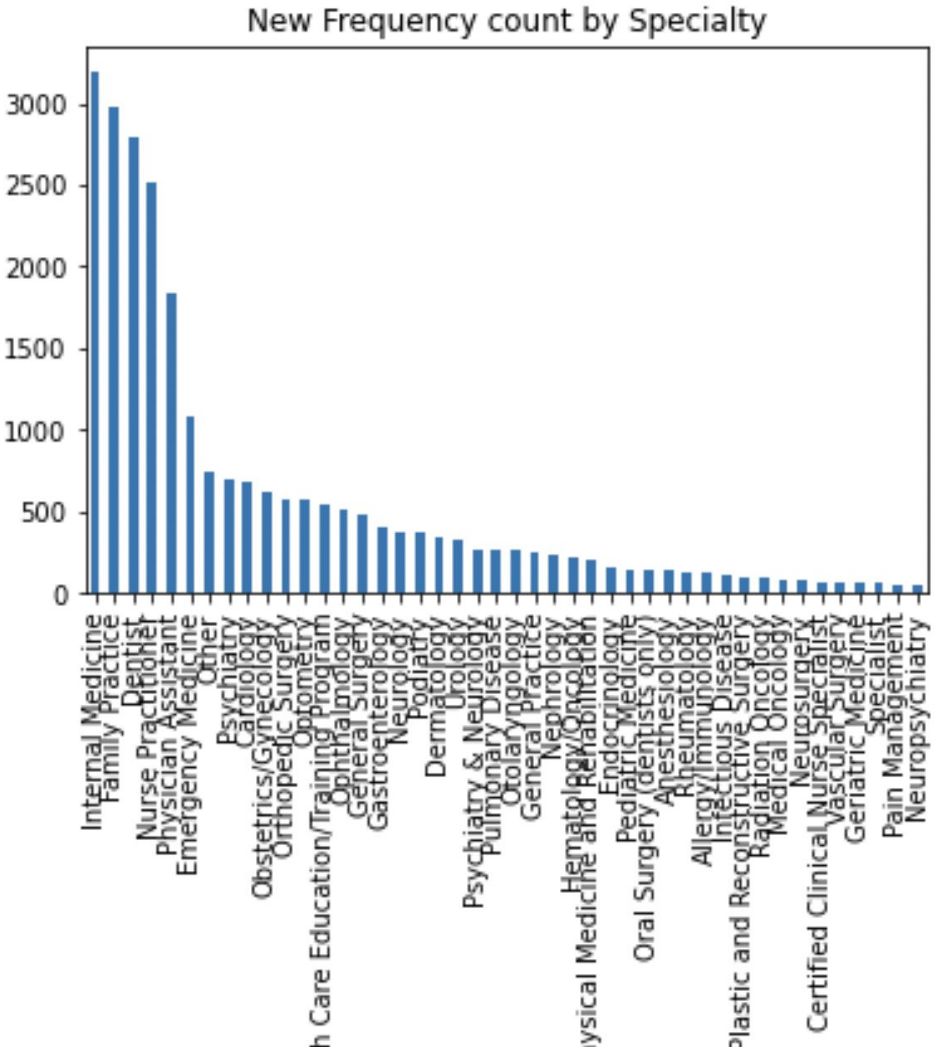
Spe

```
#now
len(d
```

```
109
```



Save features that
Frequency count>50
Save others to be “other”


New Frequency count by Specialty
3000
2500
2000
1500
1000
500
0
Internal Medicine
Family Practice
Dentist
Nurse Practitioner
Physician Assistant
Emergency Medicine
Other
Psychiatry
Cardiology
Obstetrics/Gynecology
Orthopedic Surgery
Optometry
Health Care Education/Training Program
Ophthalmology
General Surgery
Gastroenterology
Neurology
Podiatry
Dermatology
Urology
Psychiatry & Neurology
Pulmonary Disease
Otolaryngology
General Practice
Nephrology
Hematology/Oncology
Physical Medicine and Rehabilitation
Endocrinology
Pediatric Medicine
Oral Surgery (dentists only)
Anesthesiology
Rheumatology
Allergy/Immunology
Infectious Disease
Plastic and Reconstructive Surgery
Radiation Oncology
Medical Oncology
Neurosurgery
Certified Clinical Nurse Specialist
Vascular Surgery
Geriatric Medicine
Pain Management Specialist
Neuropsychiatry

One-hot encoding on Specialty

```
features_list
```

```
 'Specialty_is__Anesthesiology',
 'Specialty_is__Cardiology',
 'Specialty_is__Certified Clinical Nurse Specialist',
 'Specialty_is__Dentist',
 'Specialty_is__Dermatology',
 'Specialty_is__Emergency Medicine',
 'Specialty_is__Endocrinology',
 'Specialty_is__Family Practice',
 'Specialty_is__Gastroenterology',
 'Specialty_is__General Practice',
 'Specialty_is__General Surgery',
```

Features

Gender (0/1)

List of State_is_ (0/1)

List of Specialty_is_ (0/1)

List of Drugs (0/1)

Death_Rate_norm (float between 0 and 1)

Now There is a total of

```
len(features_df.iloc[0,:])
```

346 Features

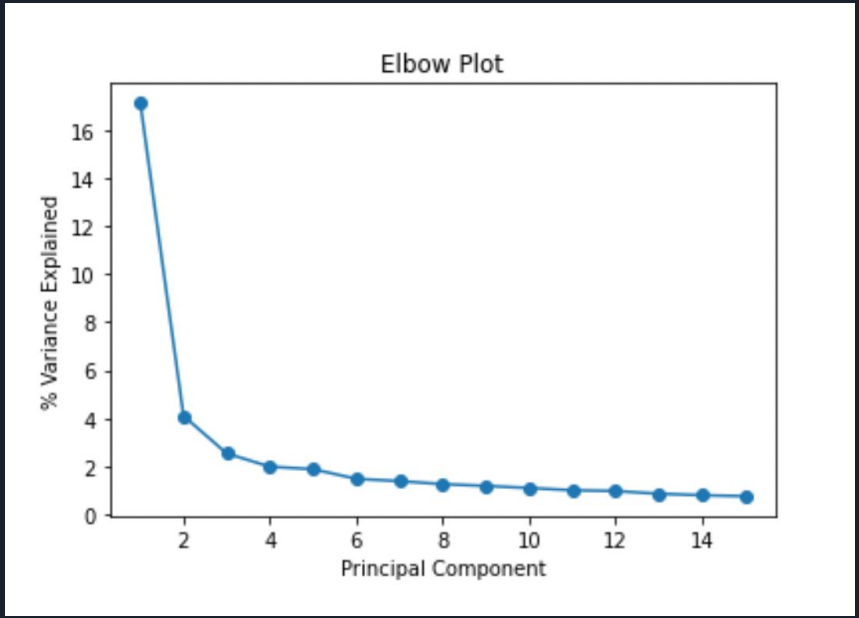
And is ready for the next step

Dimension Deduction

Dimension Deduction

Elbow Plot

- 2-3 variables

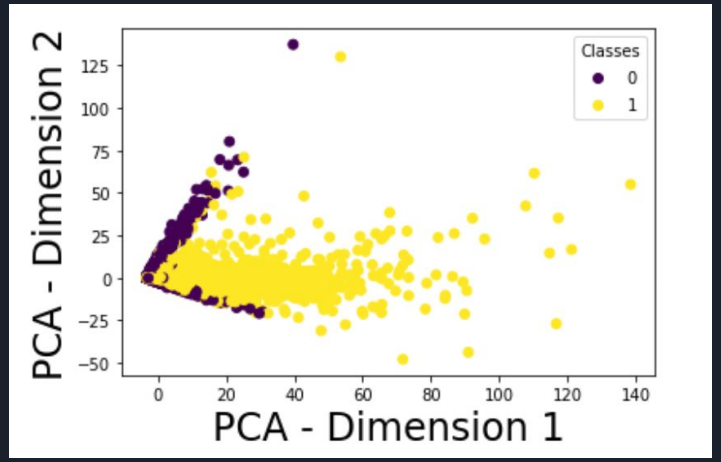


Dimension Deduction

- Visualization

2D Map of Opioid Prescriber

- Obvious trend in two classes



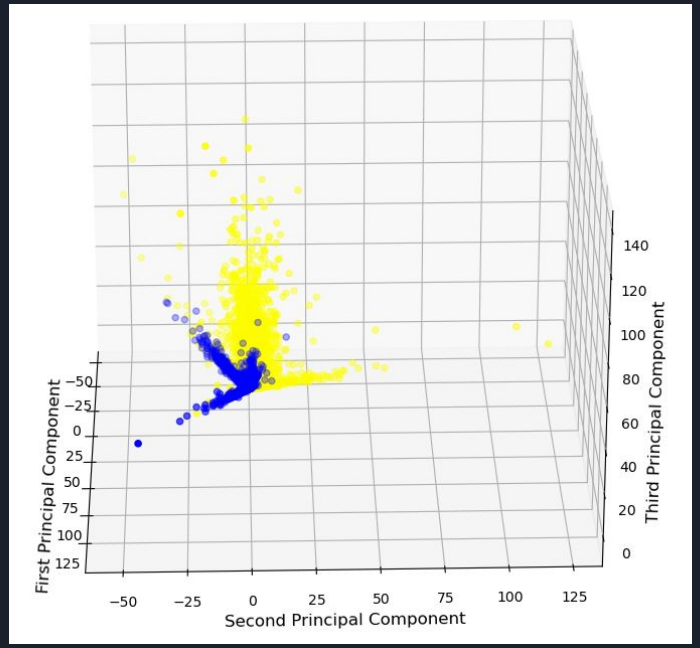
Dimension Deduction

- Visualization

Yellow: 1 Blue: 0

3D Map of Opioid Prescriber

- More obvious distribution



Dimension Deduction

- **Eliminate Dimension**

- Reduce number of attributes to 211 with 90% of the variance in the target variable explained
- Original attributes: 346

- **Good Visualization**
- **Not very significant deduction with a loss of 10%**

Methods & Models

Methods & Models

- Logistic Regression
- Decision Tree
- Random Forest (max_depth = 100)
- KNN Classifier (n_neighbors = 17)
- SVM Classifier (kernel = ‘rbf’)
- Gaussian Naive Bayes
- Classification_report of confusion matrix

❖ Focus on the prediction of 1 of Opioid Prescriber

Measures

- Target_variable: Opioid.Prescriber (frequent (1) or not (0)).
- Overall precision: Percentage of correct prediction of both 1 and 0 .
- Precision of 1: Percentage of correct prediction of frequent opioid prescribers.
- Recall: Percentage of predicted frequent opioid prescribers out of all frequent opioid prescribers in real situation.
- F1 score: harmonic mean of precision and recall

Precision of predictions

Model's name	Overall precision	Precision of 1
Logistic Regression	0.9213685474189676	0.98
Decision Tree	0.9039615846338536	0.93
Random Forest	0.9227691076430572	0.96
KNN Classifier	0.8723489395758304	0.95
SVM Classifier	0.8809523809523809	0.98
Gaussian Naive Bayes	0.9099639855942377	0.95

Recall of predicting 1s

Model's name	Recall
Logistic Regression	0.88
Decision Tree	0.91
Random Forest	0.91
KNN Classifier	0.82
SVM Classifier	0.82
Gaussian Naive Bayes	0.89

F-1 score of predicting 1s

Model's name	F-1 score
Logistic Regression	0.93
Decision Tree	0.92
Random Forest	0.93
KNN Classifier	0.88
SVM Classifier	0.89
Gaussian Naive Bayes	0.92

Key features

HYDROCODONE.ACETAMINOPHEN	0.240765
OXYCODONE.ACETAMINOPHEN	0.061513
TRAMADOL.HCL	0.060548
OXYCODONE.HCL	0.028386
ACETAMINOPHEN.CODEINE	0.020295
PREDNISONE	0.020028
GABAPENTIN	0.016958
Specialty_is__Emergency Medicine	0.014685
AMOXICILLIN	0.014468
Death_Rate_norm	0.013344

Run Time

Model's name	**Run Time (second)**
Logistic Regression	0.6
Decision Tree	0.8
Random Forest	3.9
KNN Classifier	5.4
SVM Classifier	48.6
Gaussian Naive Bayes	0.2

Conclusion

- Our goal: identify frequent opioid prescribers
- Data cleaning -> PCA -> Prediction
- Which is the optimum model? Why?
- Implications : systematic, efficient, narrow down the pool of potential overdoses
- Potential improvements

Reference

1. CDC Opioids [Online] Available from: https://www.cdc.gov/opioids/index.html

2. "U.S. Opiate Prescriptions/Overdoses". Accessed on: Apr. 15, 2022. [Online]. Available: https://www.kaggle.com/datasets/apryor6/us-opiate-prescriptions?select=prescriber-info.csv

3. "Drug Overdose Deaths in the U.S. Top 100,000 Annually," Nov. 17, 2021. Accessed on: Apr. 15, 2022. [Online]. Available: https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2021/20211117.htm

4. “Detecting Frequent Opioid Prescription” [Online]. Available: https://www.kaggle.com/code/apryor6/detecting-frequent-opioid-prescription

Thank you for listening!