Catching Medicare Fraud Providers from Sale Anomalies

09/12/20

Introduction

> "Health care fraud costs the United States tens of billions of dollars each year. Some estimates put the figure close to $100 billion a year. It is a rising threat, with national health care expenditures estimated to exceed $3 trillion in 2014."
>
> U.S. Department of Justice

We aim to tackle this data using a data-driven approach, particularly we hope to:
1. Detect patterns of fraud medicare providers.
2. Build classification models to detect these providers.

The analyses and conclusion from the current study are expected to be helpful for insurance companies, and other organizations that oversee Medicare program including federal centers such as Department of Justice and Department of Health and Human Services.

Data

1. CMS Medicare 2014 Part D datafrom Google BigQuery (23 million observations)
2. Medicare Exclusion list from the Office of Inspector General

Drugs in Medicare

A total of $ 90 billion from 83,000 providers with 2695 unique drugs were prescribed.

By total sale, Sovaldi, a medication for Hepatitis C, is the best selling. However, by amount of prescriptions, Hydrocodone, a narcotic pain relieving drug, is the most prescribed drug.

Top 10 Drugs in Medicare (By Price)
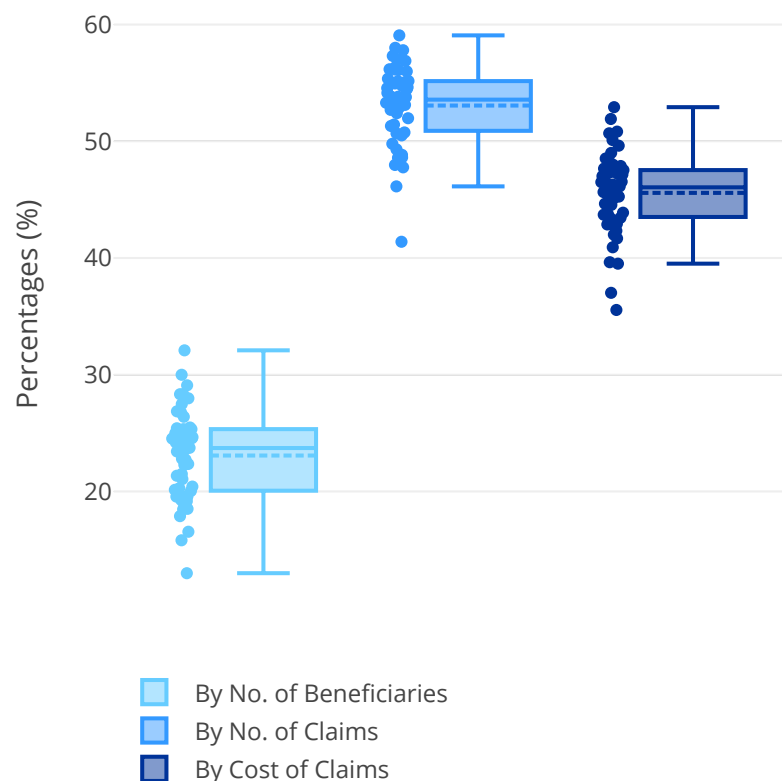
by price ▼
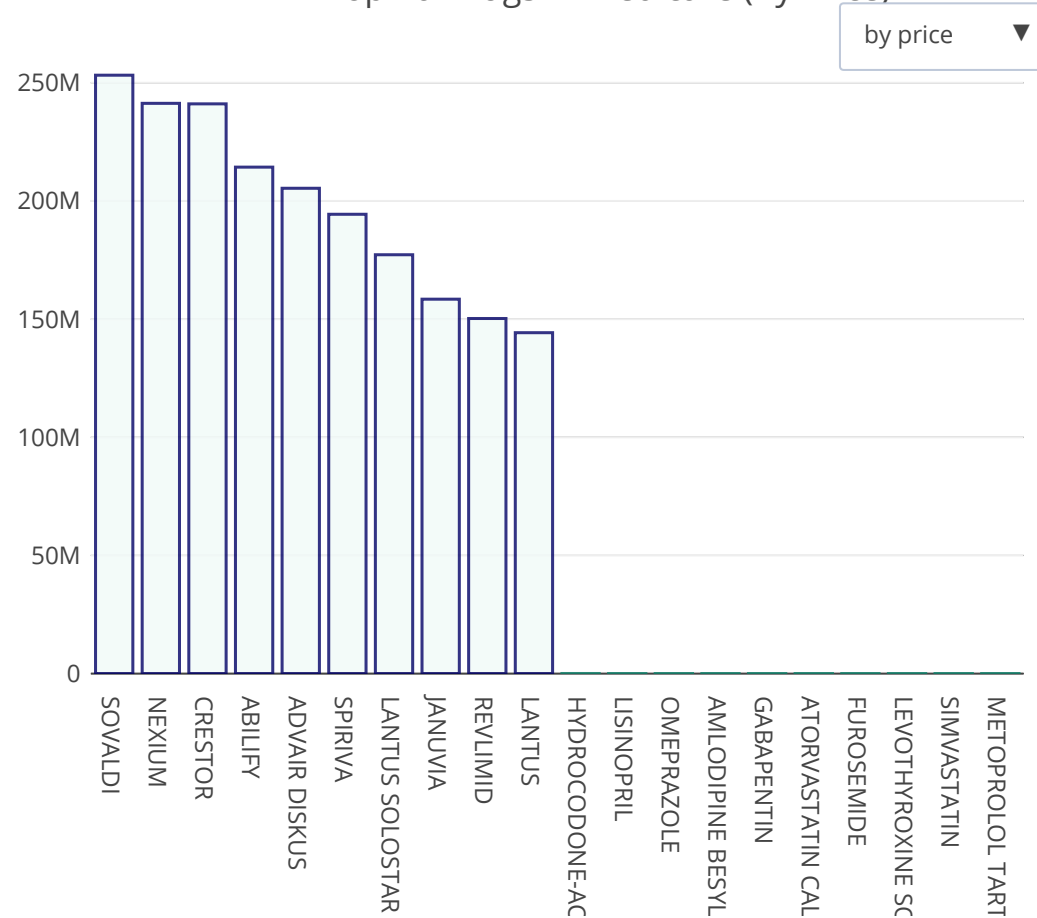
Who uses Medicare?

Across the States, only 24% of the beneficiaries are of 65+ ages (the rest are disabled beneficiaries).

However, beneficiaries with 65+ age contributed to more than 45% of the cost in Medicare prescription

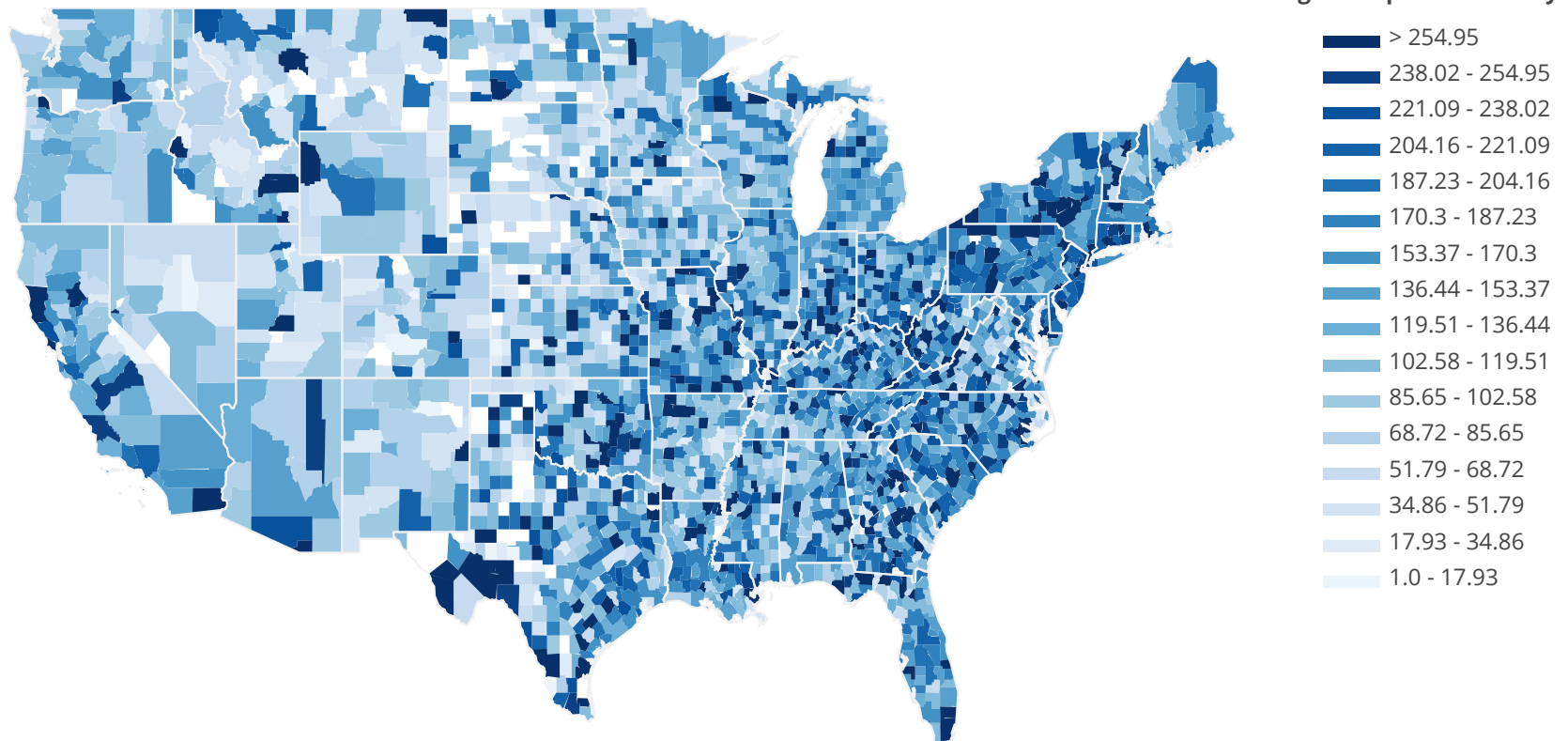
Medicare Precsription Used by Beneficiaries of 65+ Age

Percentages (%)

■ By No. of Beneficiaries
■ By No. of Claims
■ By Cost of Claims

Disparity in Medicare?

Across the States, only 24% of the beneficiaries are of 65+ ages (the rest are disabled beneficiaries).

However, beneficiaries with 65+ age contributed to more than 45% of the cost in Medicare prescription

Average Medicare Prescription Cost in the USA

Average Cost per Beneficiary

- > 254.95
- 238.02 - 254.95
- 221.09 - 238.02
- 204.16 - 221.09
- 187.23 - 204.16
- 170.3 - 187.23
- 153.37 - 170.3
- 136.44 - 153.37
- 119.51 - 136.44
- 102.58 - 119.51
- 85.65 - 102.58
- 68.72 - 85.65
- 51.79 - 68.72
- 34.86 - 51.79
- 17.93 - 34.86
- 1.0 - 17.93

Fraud providers

Fraud provider accounts to approximately 0.2%, of all providers and prescribed at least $200 million worth of drugs in the year 2014 alone.

The specialties distribution of typical providers VS fraud provider are almost the same.
This is further confirmed by Spearman's correlation, which yielded almost 0 correlations with low p-value.
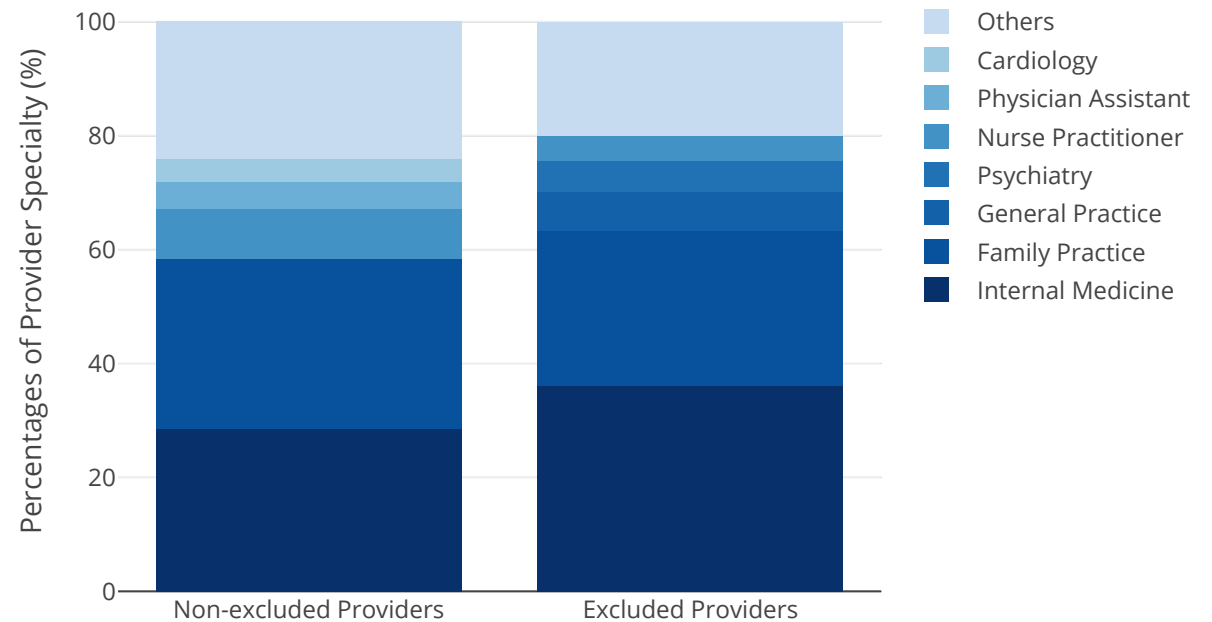
Drugs sold by fraud providers

By ranking the top 10 most prescribed drugs from only the fraud providers, more than 5 narcotics, including , Oxycontin Oxymorphone, Morphine, Oxycodone HCL, and Oxycodone-acetaminophen are on the list compared to just 1 from all providers shown previously (Hydrocodone).

Top 10 Drugs in Medicare Sold by Excluded Providers (By Price)
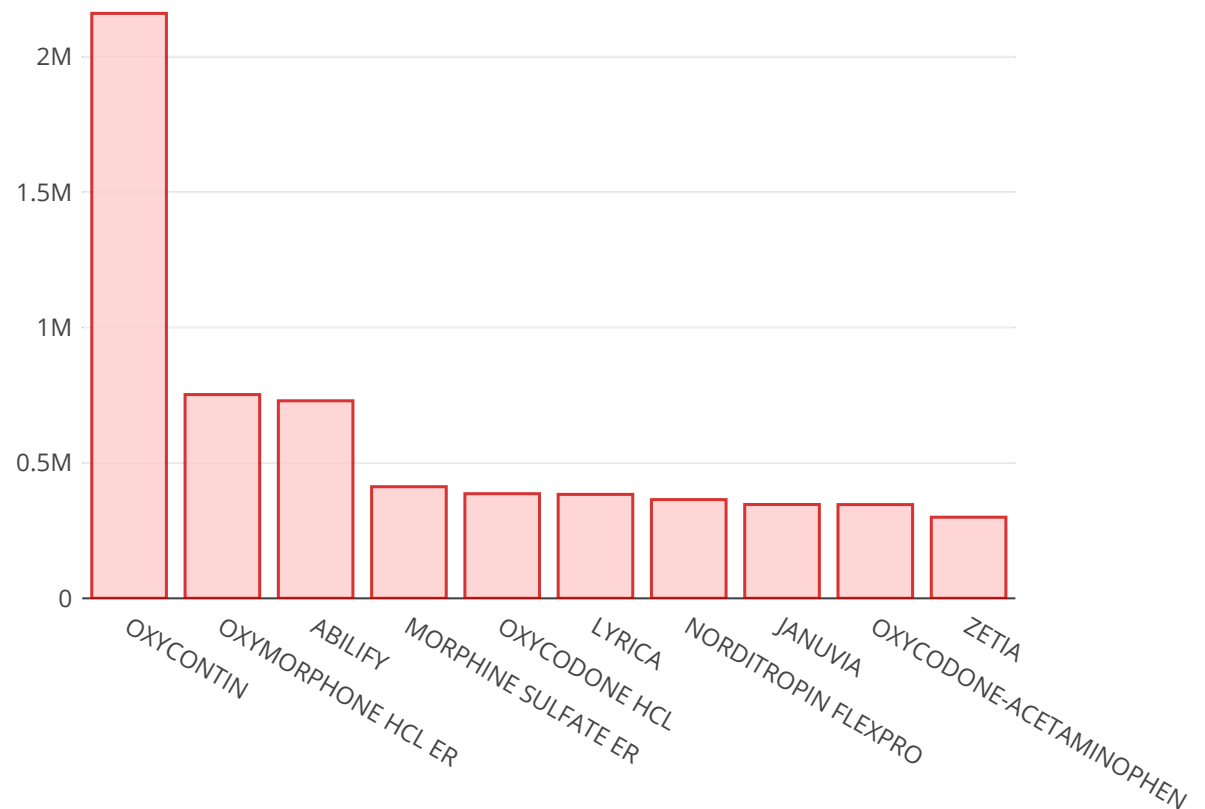
Sale differences

Typical drug providers, on average, sell almost an equal amount of non-narcotics and narcotics at similar price. In contrast, while excluded providers sell similar amount of non-narcotics at comparable price to typical providers, their narcotics sales costs more than 10x and are sold to almost 4x beneficiaries.
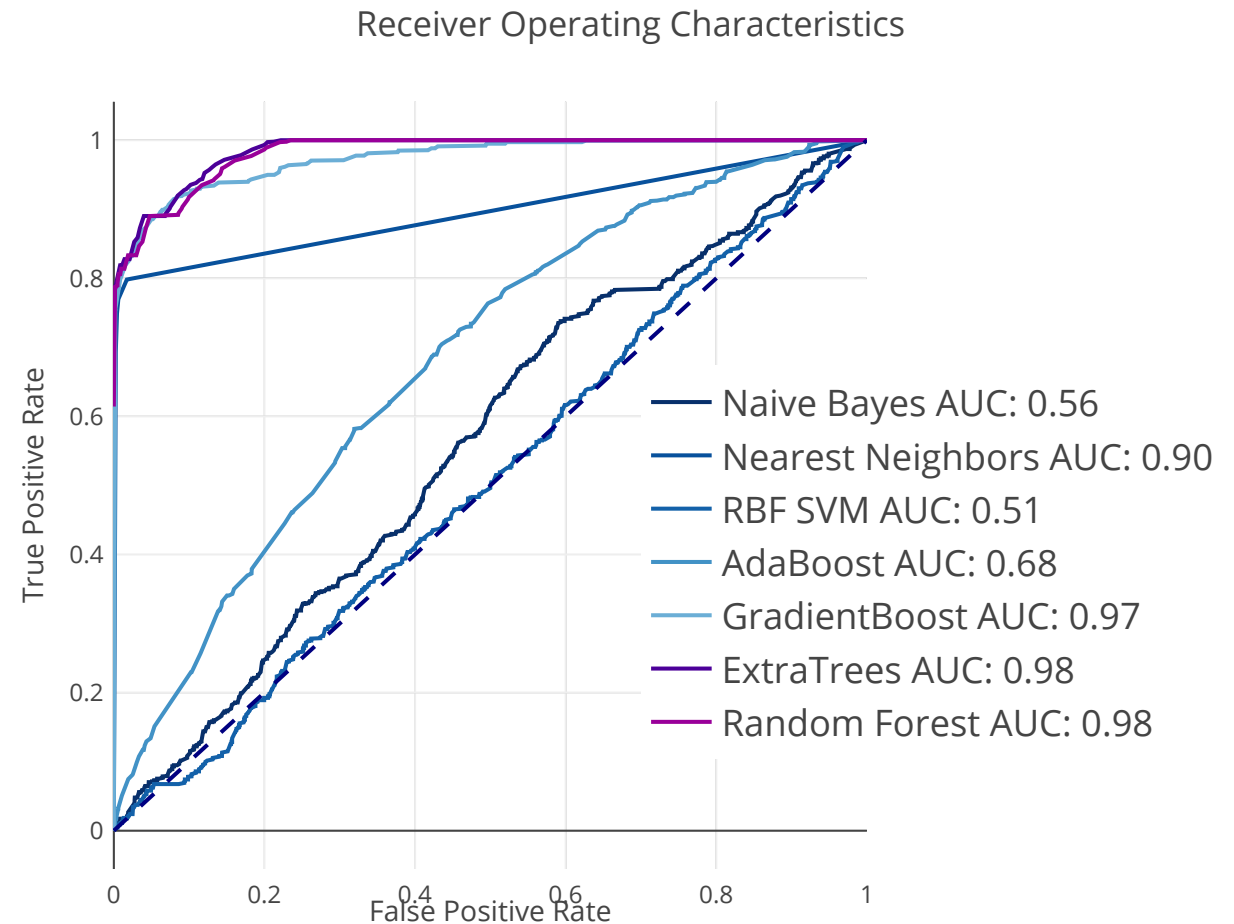

Medicare Prescription Cost and Amount by Excluded and Non-excluded Providers
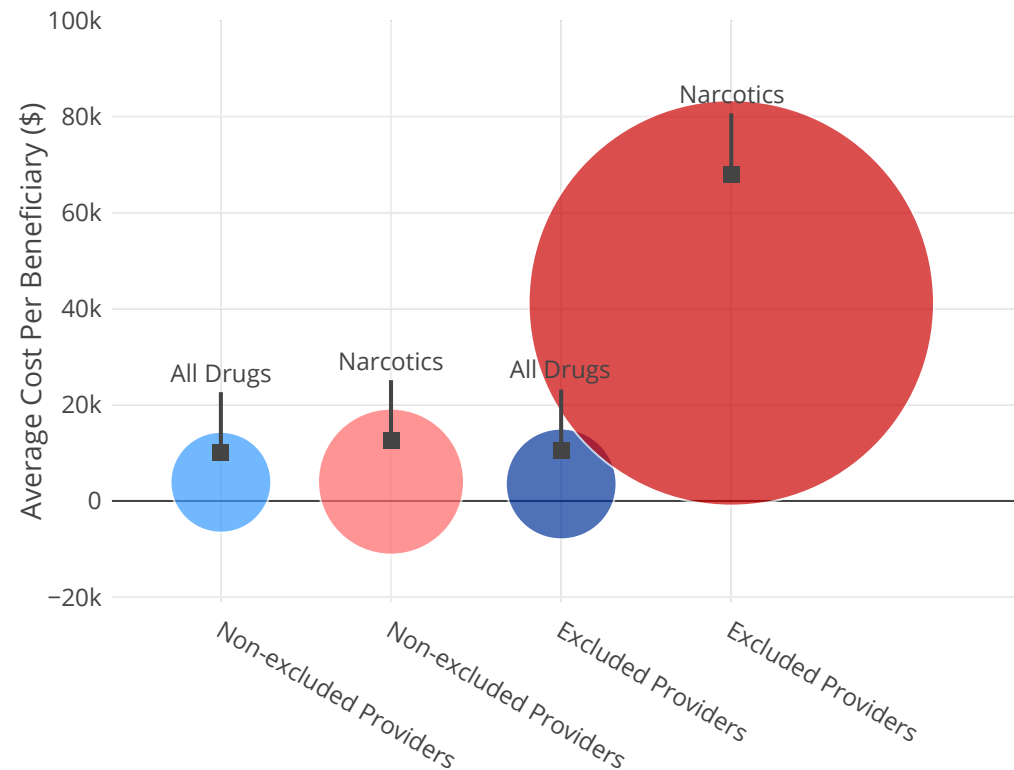
Model comparison

After feature engineering, we evaluate diffrent classification models. Typical "accuracy" cannot be used as the criteria for evaluation of the models as class imbalance (0.2% fraud provider) means that even if the model predict that all providers are not fraud, it will still show more than 99% accuracy. As such, we utilized F1 score and AUC from ROC.

Receiver Operating Characteristics

- Naive Bayes AUC: 0.56
- Nearest Neighbors AUC: 0.90
- RBF SVM AUC: 0.51
- AdaBoost AUC: 0.68
- GradientBoost AUC: 0.97
- ExtraTrees AUC: 0.98
- Random Forest AUC: 0.98

Model comparison

High F1 scores were achieved with the best at 87.5 by extra tree classifier.

Most models that yielded descent scores are based on "decision tree algorithms". This is because these trees are able to cope with class imbalance by their hierarchical structure that allows them to learn signals from both classes.

Performance of Different Models by F1 Score
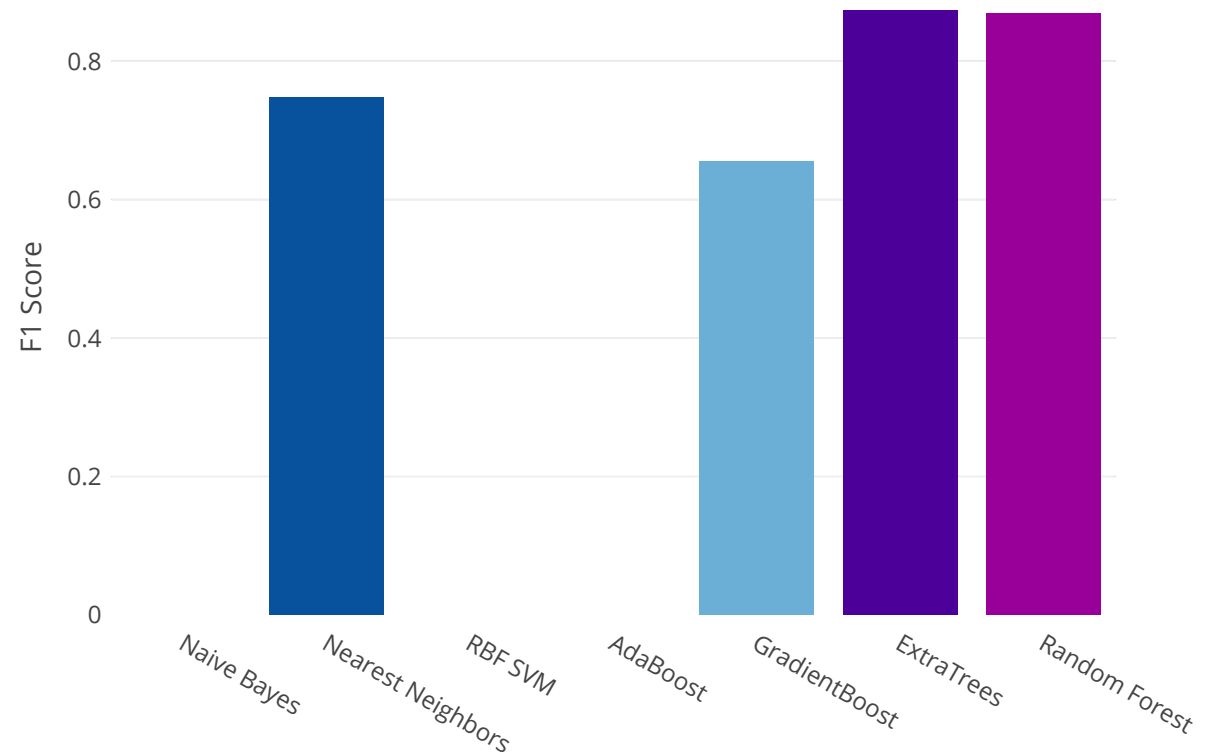
Model optimization

Extra trees model was optimized using grid search. Lastly, the model was optimized using higher number of principal components (n=3).

Performance of Random Forest with Increasing Components