Opiate/Opioid prescription analysis using machine learning

Anirudh K Muralidhar
Masters in Data Science at Indiana University
Bloomington, Indiana
anikamal@iu.edu

Arunram Sankaranarayanan
Masters in Data Science at Indiana University
Bloomington, Indiana
arunsank@umail.iu.edu

***Abstract* - The purpose of this paper is to develop a prediction model to predict if a given prescriber is a opiate prescriber or not based on their non-opiate drug prescription pattern. Also we find the top ten non-opiate drugs that influences the opiate prescription. The paper includes a model developed with decision tree embedded with a bagging classifier and logistic regression.**

***Index Terms*—Opiate, Opioid, prescriber, machine learning, decision tree, logistic regression, bagging, medical.**

Introduction

[5] Opiates are drugs that help to relieve pain by acting on the nervous system. Some common opiates include morphine, methadone, oxycodone and heroin. It can produce a feeling of euphoria that can be addictive to people which can lead to drug abuse. As a result of Opioid overdose, there has been an increase in death rate over the years in the United States.

The amount of opioid casualties in recent times in the United States were quite alarming. Some of the statistics reported by the Center for Disease and Control and Prevention(CDC) clearly suggested that opioid overdose is not something that can be taken lightly. CDC reported that opioid overdose killed at least 28000 people in 2014 which is the highest till date. On further analysis CDC found that at least half of all opioid overdose involved a prescription overdose [1]. And also a predominant amount of opioid overdose deaths were caused by cardiac arrests. This in general was a huge motivation for us to build a machine learning model to solve this problem.

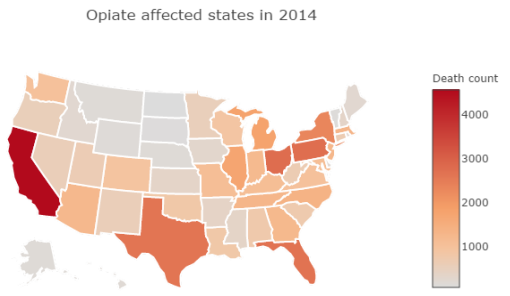

Fig 1. Opiate fatalities across the USA in 2014

There were several researches conducted to develop possible solutions for this. In this paper we use machine learning techniques to develop one. We would like to see how a non-opiate drug prescription pattern of a prescriber influences him/her to be an over the limit prescriber of opiate.

Dataset Description

The dataset has been downloaded from [2] kaggle where it originates from cms.gov. There are three files of which prescriber-info.csv displays the number of opiate and non-opiate drugs prescribed by 25,000 unique medical professionals in the United States in the year 2014 for the citizens under class D medicare. It also has some information about the professionals themselves such as their NPI number, gender, state, credentials and specialty. A doctor is labeled as opiate prescriber if they prescribe opiate drugs more than 10 times.

The overdoses.csv file contains information about the population and death count for all states in the USA. And the opioids.csv contains list of drugs that fall under the opiate category.

No	Description	Count
1	Total number of drugs	250
2	Number of Opiate drugs	11
3	Total Number of prescribers	25000
4	Opiate prescribers	14688
5	Non-opiate prescribers	10312

Table 1:Initial Opioid dataset statistics

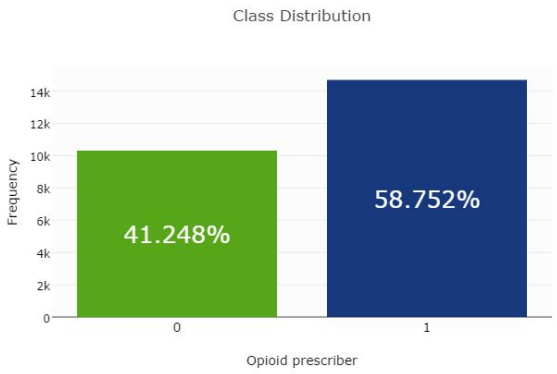

Fig 2. Class distribution of the dataset

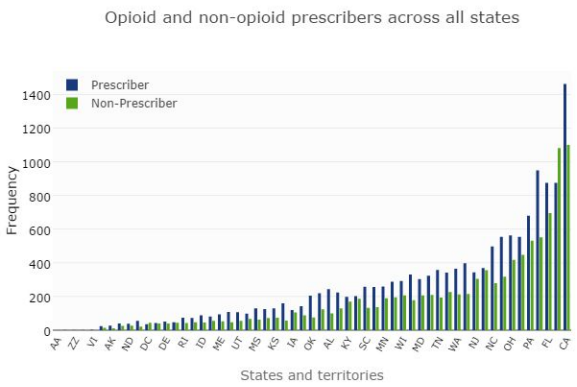

Fig 3. Opioid and non-opioid prescriber across states and territories.

Related Work

Some of the previous work relating to prediction of Opioid misuse [3] involved prediction of opioid misuse by identifying patients who showed

- Abnormality in the surveillance program conducted by researchers
- Abnormality in the urine drug screens
- Difficulties in tracking the prescriptions that are being made to patients due to their medical conditions
- Clinic staff showing a particularly poor behaviour when it comes to Opiate prescriptions

In our analysis we are particularly working on tracking and managing the opiate prescriptions made to patients to alleviate their pain by constructing a model that can predict whether a prescriber is an opiate or a non opiate prescriber. We do this by identifying the patterns in their non opiate prescriptions.

Data Preprocessing

A. *Downsampling*

From Fig 2, we could see that the class distribution is skewed towards opioid prescriber. In order to eliminate this class imbalance we perform [8] downsampling. By this way we have a much more generalized classifier.

B. *Feature space for the classifier*

We develop the feature space by removing all the opioid drugs from the data. We also remove metadata of doctors such as gender, state, NPI, credentials and specialty. Thus our feature space would be non-opioid drugs prescribed.

No	Description	Count
1	Total Number of prescribers	20624
2	Opiate prescribers	10312
3	Non-opiate prescribers	10312
4	Number of features	239

Table 2. Class distribution after preprocessing

Algorithms Used

A. *Feature selection methods*

a) *Variance threshold*

Variance threshold [4] is a feature selection method that was used on our dataset to eliminate low variance features in our feature set. This would essentially remove all features which would exhibit low variance across rows in our dataset. For this feature selection method we set different thresholds into the model by calculating the variances across different percentiles and then using them to set threshold parameters in our model.

b) *Chi square test*

[7] We use chi square test to test the dependence of each of the feature variable in the feature set to the target variable. If we find that the target variable is independent of the feature variable selected then we would remove it from our evaluation.

B. *Machine learning algorithms*

a) *Decision Tree*

Decision trees are being used for our analysis. The main reasons for using decision trees being that the decision trees

- Divide the feature space into axis parallel rectangle , thus providing an easy method for distinctly dividing the feature set as opioid or non opioid based on the input feature set
- the target function is discrete valued and decision trees work well in predicting discrete valued target functions.

b) *Logistic Regression*

- Logistic regression was the other used algorithm used in creating our model because logistic regression is better in predicting binary discrete target variable.
- Logistic regression is better for predicting target variables having a decision boundary not parallel to the axis.
- Logistic regression is less likely to over fit as it has lower variance compared to decision trees.

C *Cross validation for tuning parameters*

Five fold cross validation technique is being followed here to evaluate the performance of the created models.

The data is divided into five folds and [9] shufflesplit package is used to split the data into 5 folds and the parameter

is set to assign 40 % of the data as the test data.. The cross validation scores that were collected helped us to create a learning curve that plots the train and the test scores against different number of training examples. This helped us to give an insight about which model would work well on the given data and whether the train data overfits the model. This would help us decide about which ensemble methods would be useful if we decide to improvise the performance of the model.

D Ensemble methods

a) Bagging

Bagging is the ensemble method that has been used here to better increase the performance of the model. Bagging reduces variance and prevents the models from overfitting the data. Usually we would know that large decision trees are prone to overfit the data.

For decision trees we draw a validation curve by embedding bagging in the decision tree model. We compare the training and the cross validation scores for different bags and we see how the model performs when the number of bags are varied. This helps us in understanding whether the overfitting of training data is being avoided and the decision tree model is generalized. For logistic regression a validation curve is drawn by varying the regularization.

RESULTS AND INFERENCE

A. Without feature engineering

a) Decision Tree

The tree model was built by testing on depths ranging from 1 to 50 and using entropy as the criterion.

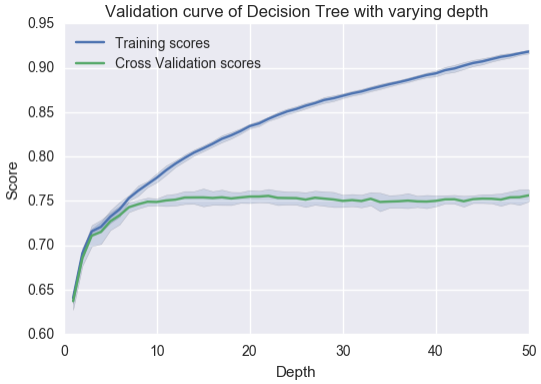

Fig 4. [10] Validation curve of decision tree across various depths

We can see that for small depths, the tree underfits the data with high bias and for high value of depths the tree overfits with high variance.

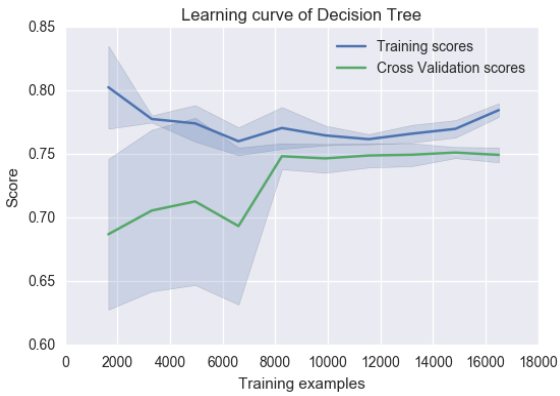

Fig 5. [10] Learning curve of decision tree for various input data with depth of 13.

It is evident that as the training example increases the cross validation score improves. Also, we could see that the learning almost saturates when the training example reaches around 10000 samples.

Accuracy: 75.3%

	Precision	Recall	F1-Score	Support
0	0.7	0.87	0.78	10312
1	0.83	0.63	0.72	10312
avg/total	0.77	0.75	0.75	20624

Table 3. Performance parameters for decision tree

	Predict negative	Predict positive
Actual negative	8968	1324
Actual positive	3764	6548

Table 4. Confusion matrix of the decision tree

Threshold	2	1	0
FPR	0	0.12	1
TPR	0	0.63	1

Table 5. False Positive Rate and True Positive Rate of the model

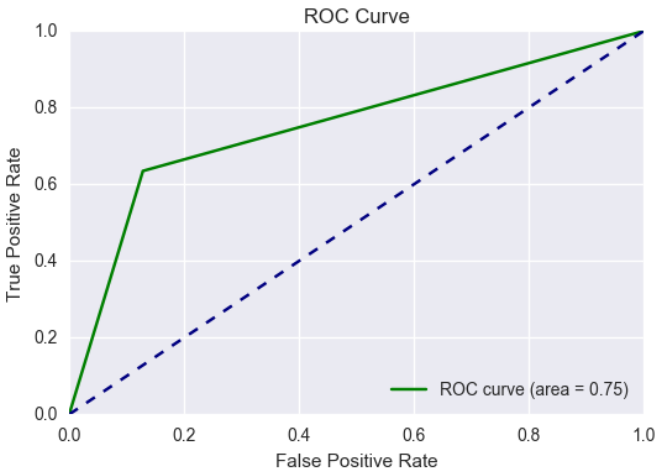

Fig 6. ROC curve of the decision tree

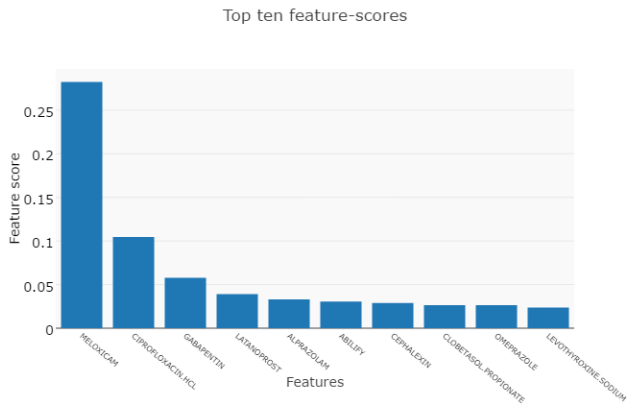

Fig 7. Top 10 features that influence a doctor to be an opioid prescriber

The top five features are as follows

- [6] Meloxicam(0.27) is a drug used to treat pain related to arthritis.
- [6] Ciprofloxacin(0.11)is a drug used to fight bacterial infections.
- [6] Gabapentin(0.06) is used to treat nerve pain caused by herpes virus.
- [6] Latanoprost(0.04) is used to treat increased pressure in eyes.
- [6] Alprazolam(0.035) is used to treat panic disorder and anxiety.

We could see that drugs such as Meloxicam and Gabapentin are used for pain treatment as well, which is kind of what opioid does.

b) Bagging classifier

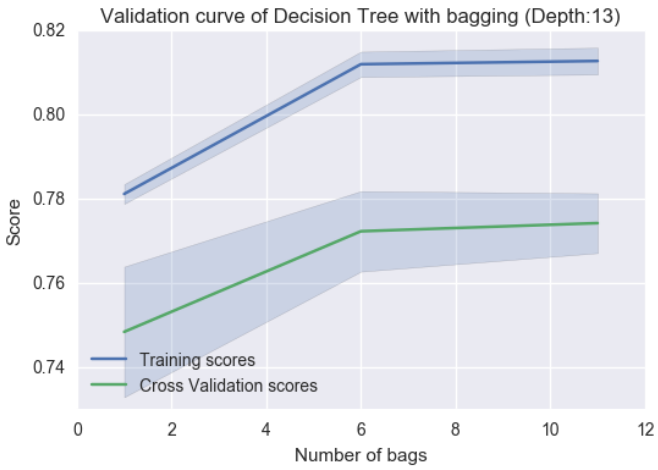

Fig 8. [10] Validation curve of decision tree(depth:13) tested with several bags

We could see the effect of bagging from this graph, as the number of bags increases the training and cross validation score improves. Let's see how the metrics have changed. The results below are published with number of bags equal to six.

Accuracy: 77.28%

	Precision	Recall	F1-Score	Support
0	0.72	0.88	0.80	10312
1	0.85	0.66	0.74	10312
avg/total	0.79	0.77	0.77	20624

Table 6. Performance parameters of decision tree with bagging

	Predict negative	Predict positive
Actual negative	9125	1187
Actual positive	3497	6815

Table 7. Confusion matrix of decision tree with bagging

Threshold	2	1	0
FPR	0	0.11	1
TPR	0	0.66	1

Table 8. False Positive Rate and True Positive Rate of the model

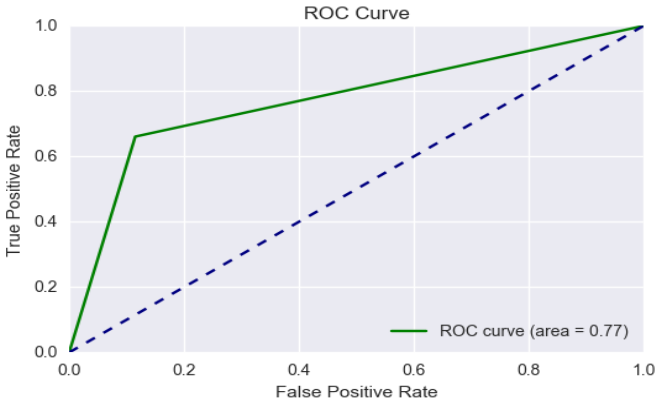

Fig 9. ROC of the decision tree with bagging

Comparing the results, we could see that the bagging has definitely improved the performance of the decision tree.

c) Logistic Regression

The logistic regression model has been developed with the various inverse regularization strength values.

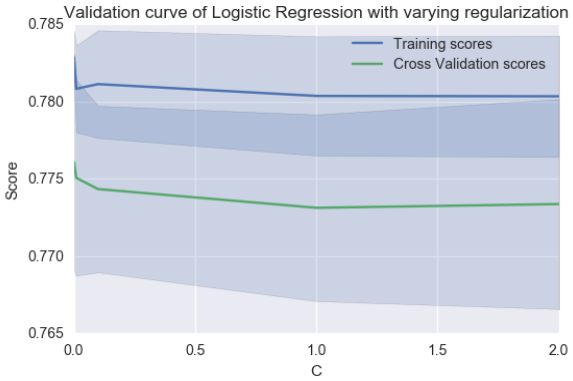

Fig .10. [10] Validation curve of logistic regression with tuning the parameter C.

From the above we could see that changing the parameter C has no major effect in the scores of logistic regression. But the best value of C was found to be 0.001.

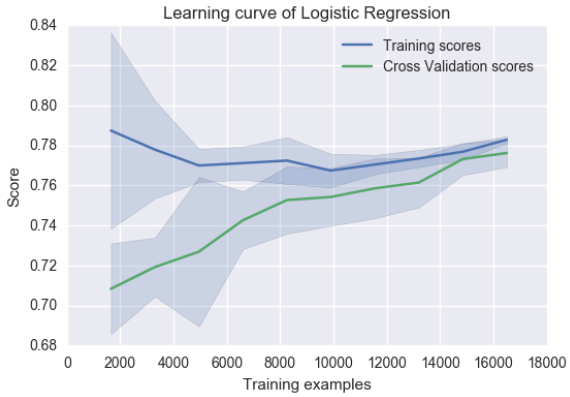

Fig 11. [10] Learning curve of logistic regression with C equal to 0.001

From the learning curve, we can infer that as the training example increases logistic regression becomes a better model.

Accuracy: 77.6%

	Precision	Recall	F1-Score	Support
0	0.73	0.88	0.80	10312
1	0.85	0.67	0.75	10312
avg/total	0.79	0.78	0.77	20624

Table 9 .Performance parameters of logistic regression

	Predict negative	Predict positive
Actual negative	9057	1255
Actual positive	3363	6949

Table 10. Confusion matrix of logistic regression

Threshold	2	1	0
FPR	0	0.12	1
TPR	0	0.67	1

Table 11. False Positive Rate and True Positive Rate of the model

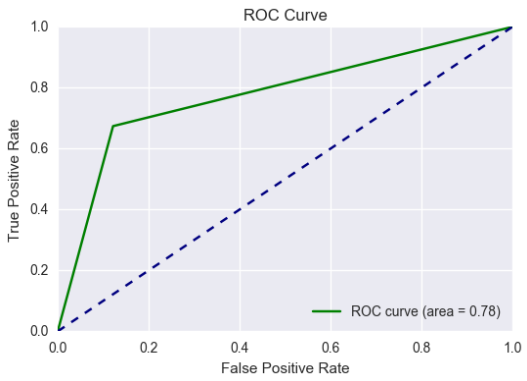

Fig 12. ROC curve of logistic regression

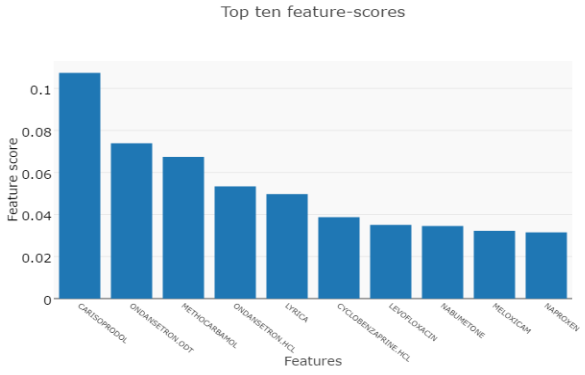

Fig 13. Feature importance based on logistic regression

The top five features which influences a doctor to be opioid prescriber are as follows

- [6] Carisoprodol(0.12) is a muscle relaxer that works by blocking pain signals between brain and nerves.
- [6] Ondansetron.ODT(0.07) is used to prevent nausea and vomiting.
- [6] Methocarbamol(0.063) is also a muscle relaxer that blocks pain signals between brain and nerves.
- [6] Ondansetron.HCL(0.056) is similar to Ondansetron.ODT.
- [6] Lyrica(0.052) can block chemicals that sends pain signals to the nerves.

From this we can infer that top features has drugs which are used to treat pain that are related to nerves.

d) Comparison of decision tree and logistic regression

Though we could see similar values in the metrics obtained from decision tree and logistic regression there are few notable differences as follows.

- From the learning curve we can see that decision tree saturates at around 10000 points, but logistic regression keeps getting better as the training example increases.
- There is a difference in the top features that make an influence on the opioid prescriber and also even the score assigned also varies.

These differences might pop because of the difference in the decision boundary that these classifiers learns.

B) With feature engineering

a) Decision Tree

The first feature engineering is done with variance threshold method which gives the following results.

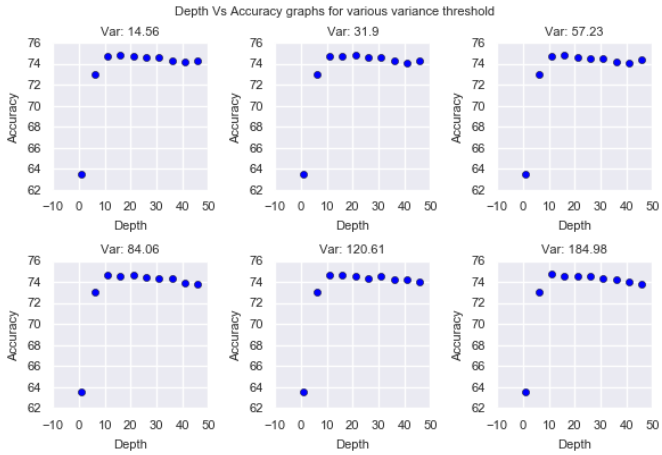

Fig 14. Feature engineering with variance threshold across various depths

With this method, we found out that decision tree has the best score with depth 11 and variance threshold of 184.9. So after this step, we reduce the number of features to 119 from 239.

Followed by this, we do chi-square test to select the top k features.

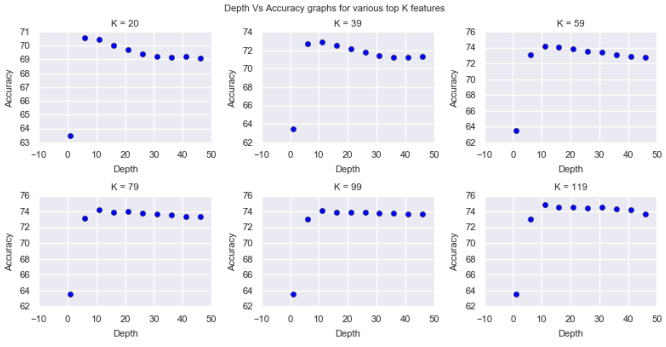

Fig 15. Scores of the model when tested with several top K features and various depths.

After this process we found that the model gives best result with depth 11 and top 119 features. Going with these values, we obtain the following results.

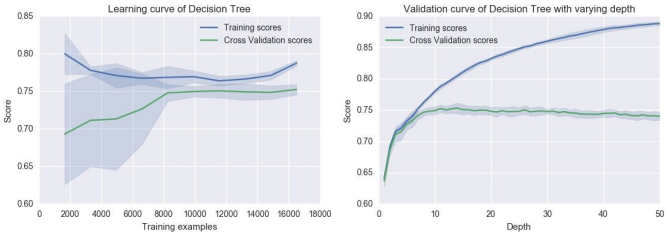

Fig 16.. Learning and validation curve of the decision tree. Depth of 11 used in learning curve.

From learning curve we could see that the model saturates after 10000 train examples like the one developed without feature engineering. From the validation curve, we can see the tree under fits for lower depths and overfits for higher depths.

Accuracy: 75.2%

	Precision	Recall	F1-Score	Support
0	0.71	0.87	0.78	10312
1	0.83	0.64	0.72	10312
avg/total	0.77	0.75	0.75	20624

Table 12. Performance parameters of decision tree with feature engineering

	Predict negative	Predict positive
Actual negative	8924	1388
Actual positive	3721	6591

Table 13. Confusion matrix of decision tree with feature engineering

Threshold	2	1	0
FPR	0	0.13	1
TPR	0	0.64	1

Table 14. False Positive Rate and True Positive Rate of the model

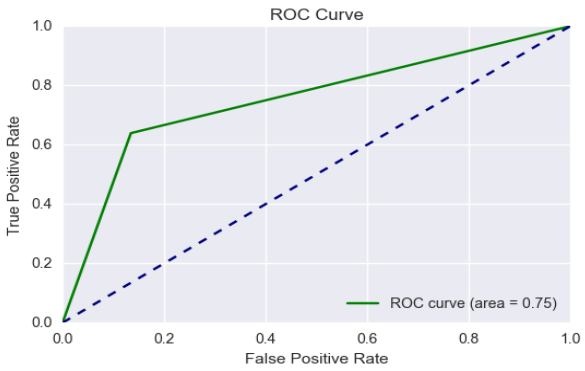

Fig 17. ROC curve of decision tree with feature engineering

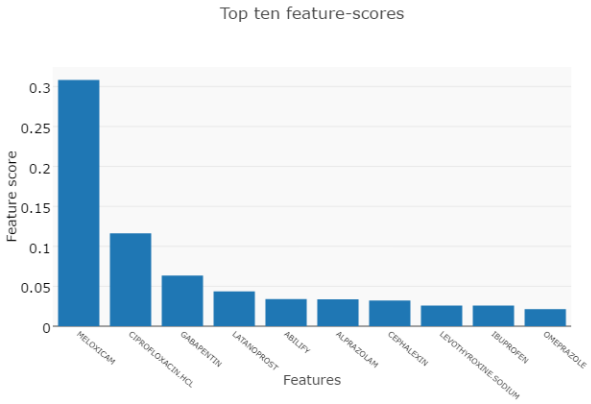

Fig 18. Top features and scores based on decision tree with feature engineering

We could see the above features looks similar to the one which we got without feature engineering, we just have one change in top 5, abilify which is a drug used to treat bipolar I disorder is that change.

b) Effect of bagging classifier on decision tree

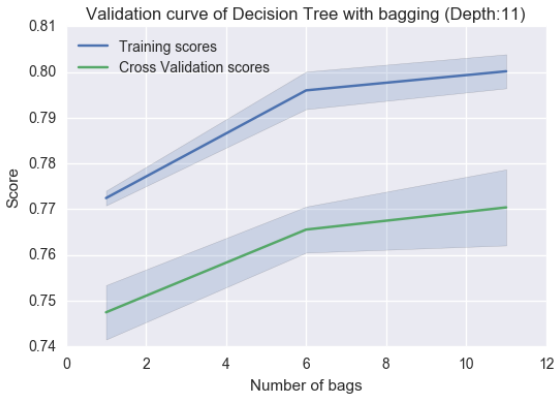

Fig 19..[10] Effect of bagging on the decision tree

We know that a complex decision trees can have high variance. A bagging classifier helps to reduce the variance which can be seen in the improvement of scores.

c) Logistic Regression

Applying the variance threshold as the first feature engineering step, we have the following results.

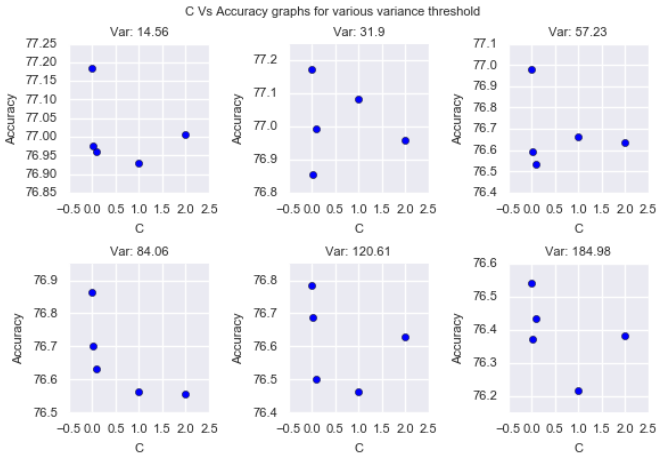

Fig 20. Feature engineering with variance threshold tested on various values of C.

From this we found out that variance threshold of 14.56 and C of 0.01 gives us best result. Based on this we reduce the number of features to 236.

By applying chi-square test to get top-k features we have the following results.

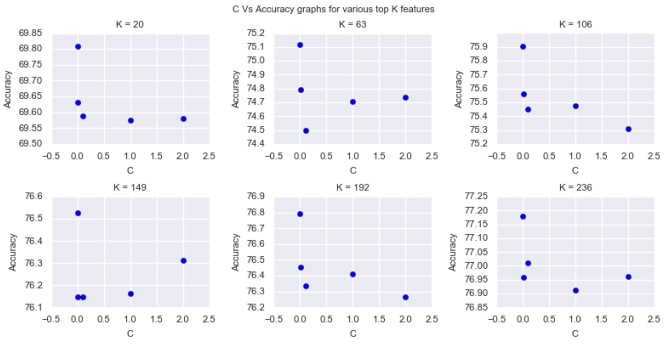

Fig 21. Feature engineering using chi-square test for several top k features tested across various C values.

From this we could obtain maximum results when we go with 236 features itself.

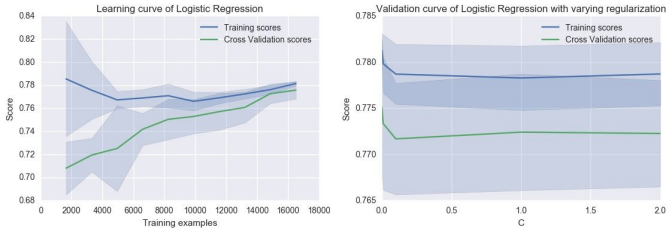

Fig 22. [10] Learning and validation curve of logistic regression. Learning curve uses the C value as 0.001.

From the validation curve we can see that logistic regression performs the same for almost all C values. But for very small values it does well.

The learning curve shows that logistic regression becomes a better model as the training data increases.

Accuracy: 77.5%

	Precision	Recall	F1-Score	Support
0	0.73	0.88	0.80	10312
1	0.85	0.67	0.75	10312
avg/total	0.79	0.78	0.77	20624

Table 15. Performance parameters of logistic regression with feature engineering

	Predict negative	Predict positive
Actual negative	9060	1252
Actual positive	3387	6925

Table 16. Confusion matrix of logistic regression with feature engineering

Threshold	2	1	0
FPR	0	0.12	1
TPR	0	0.67	1

Table 17. False Positive Rate and True Positive Rate of the model

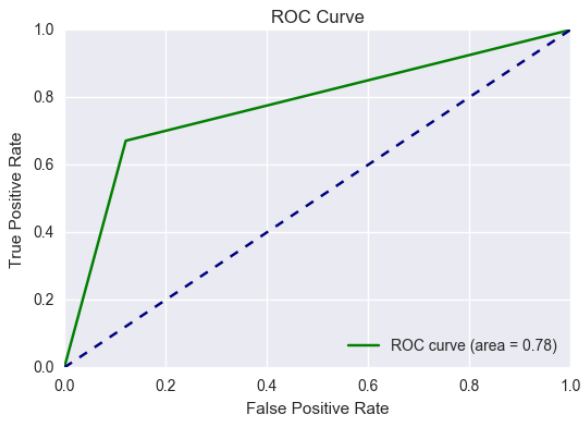

Fig 23. ROC curve of logistic regression

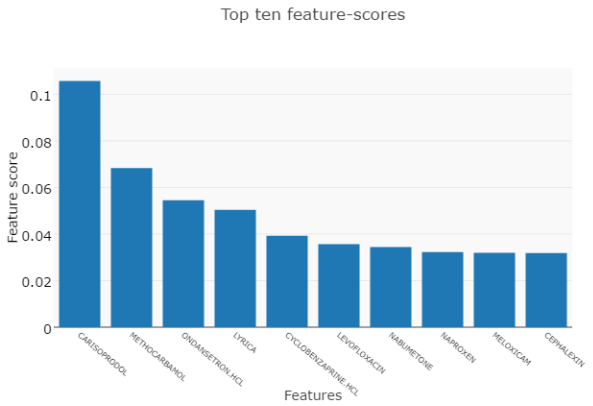

Fig 24. Top ten featured obtained with logistic regression

We could see similar features which showed with logistic regression without feature engineering.

C. Comparison of results with and without feature engineering

From the results above we can say that feature engineering did not have much impact in the improving the performance of the model. This might be because of following

- The dataset is heavily sparse.
- Even after feature engineering we have more than 100 features.

D. Average confusion matrix

Decision Tree

	Predict negative	Predict positive
Actual negative	8964	1356
Actual positive	3742	6570

Table 18. Average confusion matrix for decision tree

Logistic regression

	Predict negative	Predict positive
Actual negative	9059	1254
Actual positive	3375	6937

Table 19. Average confusion matrix of logistic regression

Overall average confusion matrix

	Predict negative	Predict positive
Actual negative	9012	1305
Actual positive	3556	6754

Table 20 Average confusion matrix

FUTURE WORK

Some possible advancements and future works that could be done are

- Training the algorithms with even more number of points might improve the performance results of logistic regression
- Knowing the properties of these non opioid drugs can help us find opioid substitutes amongst them.
- Other convoluted algorithms like SVM and neural networks can be implemented to verify if they can produce better results.

INFERENCE & CONCLUSION

This project developed few predictive models based on algorithms such as decision tree and logistic regression and fine tuned its parameters based on validation test data. The main inferences that we can make are as follows

- Referring fig 4 and 16, we can say decision tree underfits for low depth values and overfits the data as depth increases.
- From Fig 11 and 22, we can say logistic regression does a better job as training examples increases.
- Bagging does improve the performance of decision tree.
- Feature engineering might not have much impact because of the data being so sparse.

To conclude, we could make use of these models to identify doctors who generally prescribe high opiates. We can also analyze the situations in which a doctor prescribes opiate drugs and find other alternate drugs for that. By this way, we could reduce the usage of opiate and in turn reducing the fatalities caused by it.

Acknowledgment

This project is done for the course Applied Machine Learning. And we would like to thank our Professor Sriraam Natarajan for teaching this course and the AIs for their inputs.

References

[1] https://www.cdc.gov/drugoverdose/

[2] https://www.kaggle.com/apryor6/us-opiate-prescriptions

[3] **A New Tool for Prediction of Opioid Misuse -** Mark L Gostine, MD, mlgostine@iserv.net[1], Fred N Davis, MD[1], Rebecca J Risko, BSN RN[2], David Gostine, BS[3], Ajay D Wasan, MD MSc[4], (1) Michigan Pain Consultants, Grand Rapids, MI, (2) ProCare Systems, Grand Haven, MI, (3) Georgetown University School of Medicine, Washington, DC, (4) University of Pittsburg, Pittsburgh, PA

[4] http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.VarianceThreshold.html

[5] http://drugabuse.com/library/opiate-abuse/

[6] https://www.drugs.com/

[7] https://en.wikipedia.org/wiki/Chi-squared_test

[8] http://www.simafore.com/blog/handling-unbalanced-data-machine-learning-models

[9] http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.ShuffleSplit.html

[10] http://scikit-learn.org/stable/modules/learning_curve.html