

CONSUMER COMPLAINT ANALYSIS

Chris Michael | Galvanize Data Science Immersive
Seattle Cohort 4

INTRODUCTION

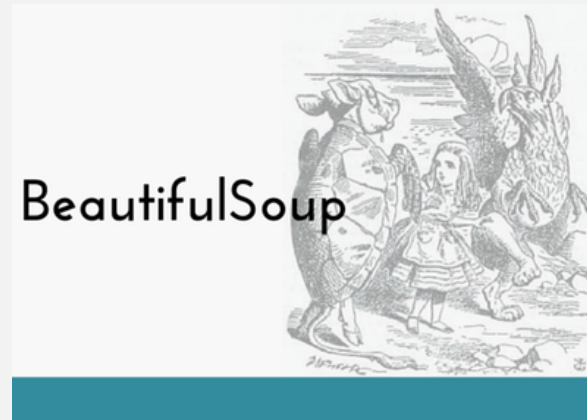
- Mechanical Engineering (BSME)/Electrophysicist (MSEE)
- Lead Landing Gear Analyst – 787, 777X
- I love telling a story with data but needed better tools!

OBJECTIVE

- Can the resolution of a consumer complaint be predicted based on information contained in text data?

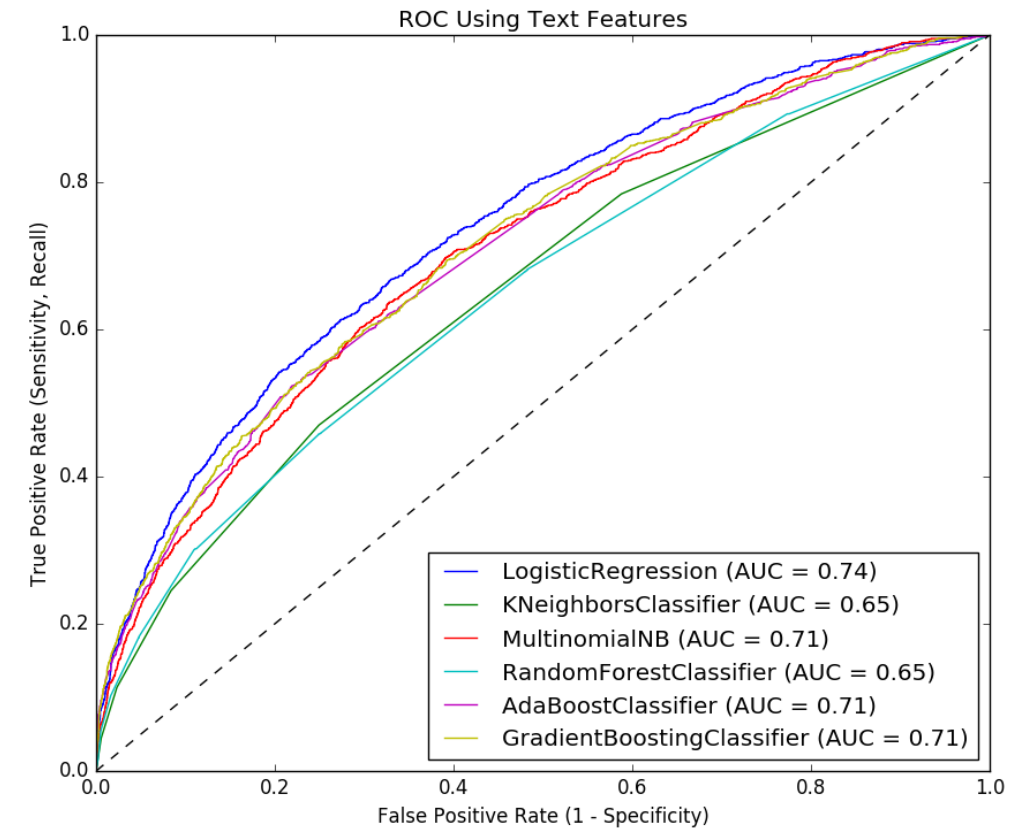
DATA

- Data from the Consumer Financial Protection Bureau (CFPB)
- Using only closed cases with text narrative
- Data is imbalanced (80% “No Relief”, 20% “Relief Provided”) - Addressed with weighting in SKLearn



MODELING/RESULTS

Model	Accuracy	Precision	Recall	F1	AUC
Using Non-Text Features					
Logistic Regression	53.0				
Using Text Features					
Logistic Regression	82.2	65.0	21.3	32.1	76.1
KNeighbors	37.5	21.0	78.4	33.2	55.8
Multinomial Naïve Bayes	81.5	65.7	13.5	22.4	74.3
Random Forest	81.1	59.0	14.3	23.0	68.3
AdaBoost	82.0	65.2	18.9	29.3	73.3
Gradient Boosting	82.1	77.8	13.4	22.9	73.7



NOTE – Table shows data from full data set, ROC Plot shows data from 20000 sample dataset

NEXT STEPS

Complaint prediction provides insight to:

- Manage costs
- Position resources
- Modeling with Word2Vec
- Analyze complaint processing time
- Predict results for specific product
- Analyze additional text features in data set.

Chris Michael

GitHub: https://github.com/ChrisMichael11/Consumer_Complaint_Project

LinkedIn: <https://www.linkedin.com/in/christophermmichael>

Email: christopher.m.michael@gmail.com