CAPSTONE 3: INCREASE CUSTOMER SATISFACTION WITH MACHINE LEARNING

JOE BOARDMAN

PROBLEM

- Complaint classification at big banks to improve customer satisfaction

RESULT

- Effective multi-class classification model achieving a 0.94 AUC score
- Topic modeling to uncover a more accurate root cause for complaint
  - Customer Service issues, etc

DATA

- Consumer Financial Protection Bureau (CFPB) complaint database
- Focused only on 4 large American banks
  - JP Morgan Chase
  - Bank of America
  - Wells Fargo
  - Citibank
- 600k records

	Product	Issue	complaint_text	Company	company_response	disputed	complaint_ID	sentences	words	special_chars	preprocessed_complaint
0	Other financial service	Fraud or scam	Seller scammed me over XXXX. Signed a contract...	JPMORGAN CHASE & CO.	Closed with explanation	0	3529560	3	61	0	seller scammed sign contract service send paym...
1	Credit card or prepaid card	Billing statement	Macys is charging me {$2.00} per month on acco...	CITIBANK, N.A.	Closed with monetary relief	0	2271267	4	78	0	macys charge 2 per month account zero balance ...
2	Credit card or prepaid card	Advertising and marketing, including promotion...	In XXXX of 2017 when i reviewed my credit ...	CITIBANK, N.A.	Closed with explanation	0	2494118	4	105	0	2017 review credit report notice best buy cred...
3	Bank account or service	Account opening, closing, or management	in short they closed my account and are withho...	JPMORGAN CHASE & CO.	Closed with explanation	0	2266442	14	316	0	short close account withhold money account cha...
4	Credit card or prepaid card	Problem with a purchase shown on your statement	Bank of America is one of the worst companies ...	BANK OF AMERICA, NATIONAL ASSOCIATION	Closed with explanation	0	2658376	10	180	1	bank america one worst company deal first no w...

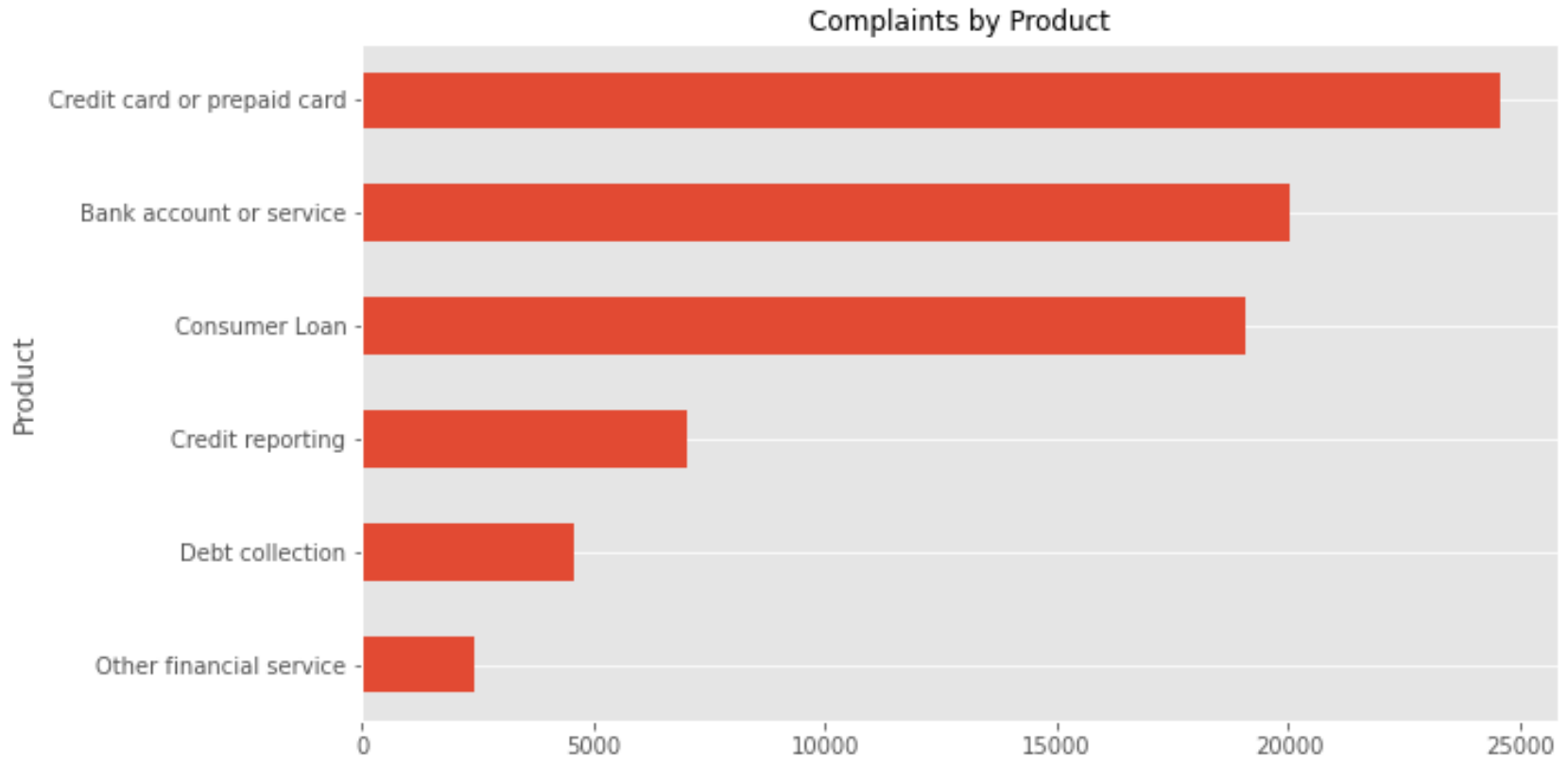
DATA CLEANING

- Filtered records without written complaint
- Sampled from full dataset to only include 4 banks
  - Left with 77k rows
- Consolidated products to solve redundancy
- Checked for duplicates and removed if necessary

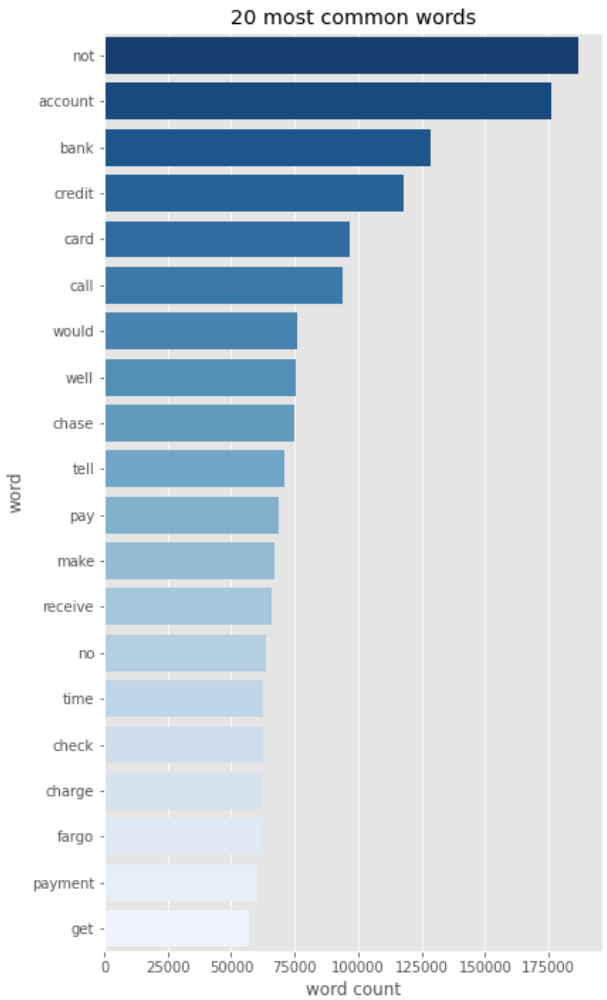
TEXT PREPROCESSING

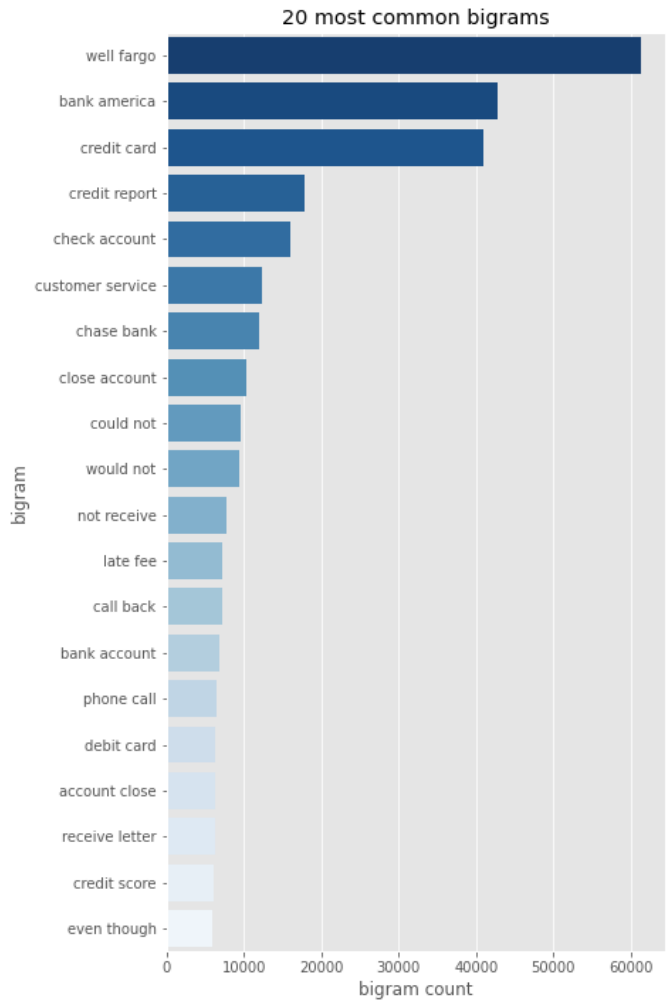
- Word tokenization
- Lemmatization
- Removal of stop words
- Convert back to a string

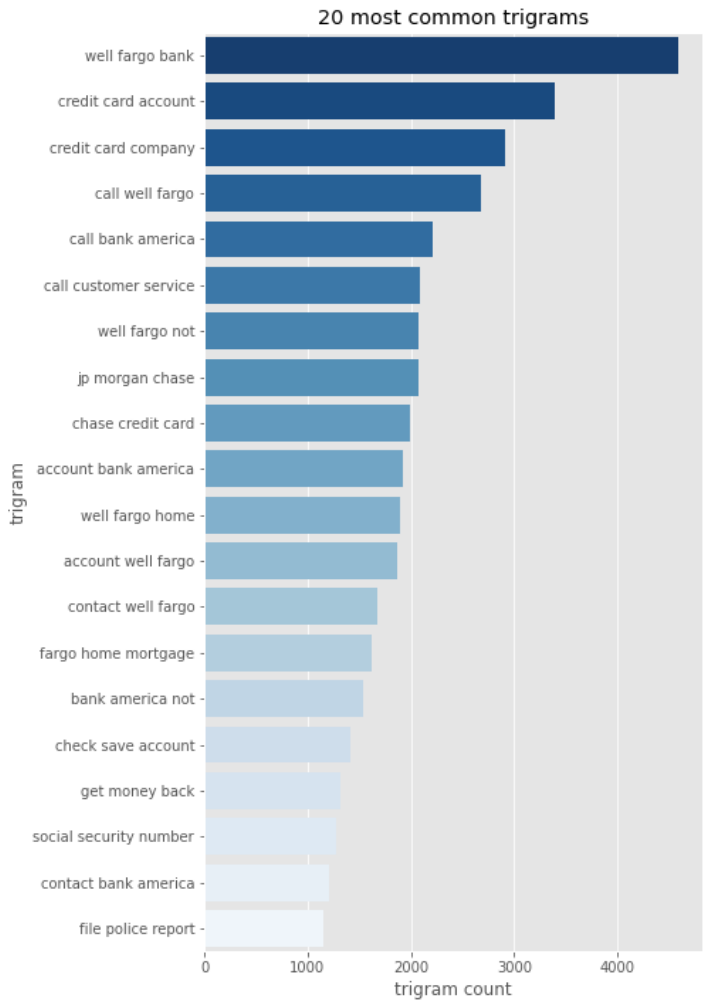
EXPLORATORY DATA ANALYSIS



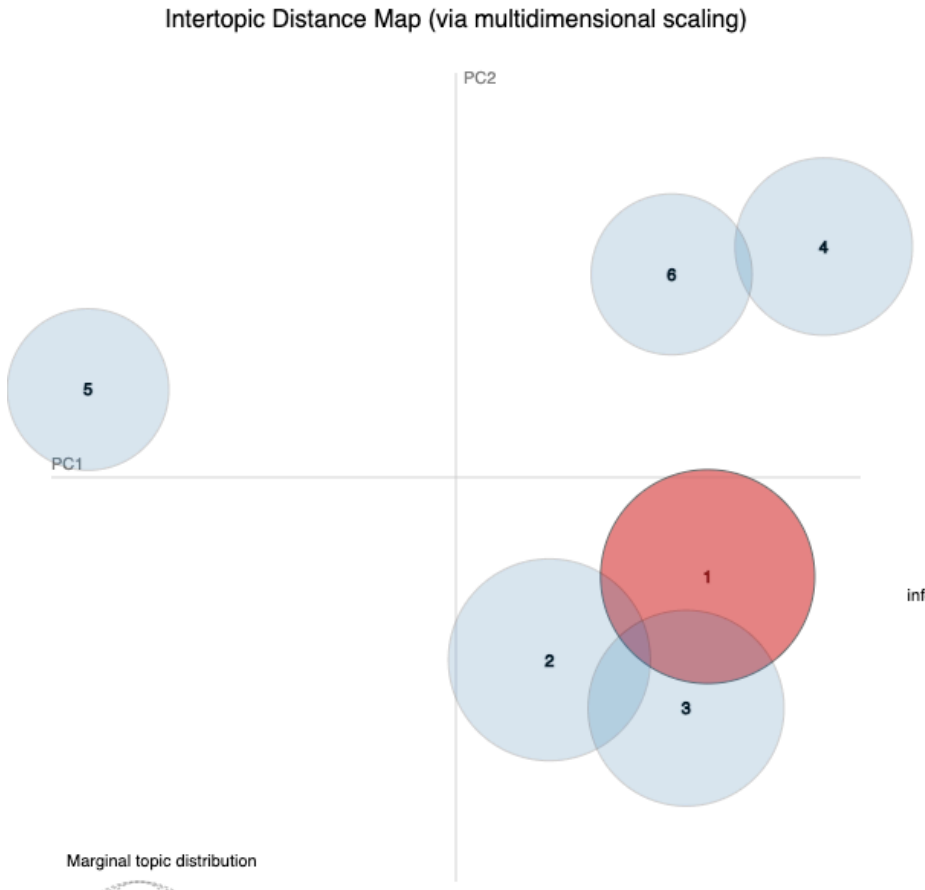
EXPLORATORY DATA ANALYSIS

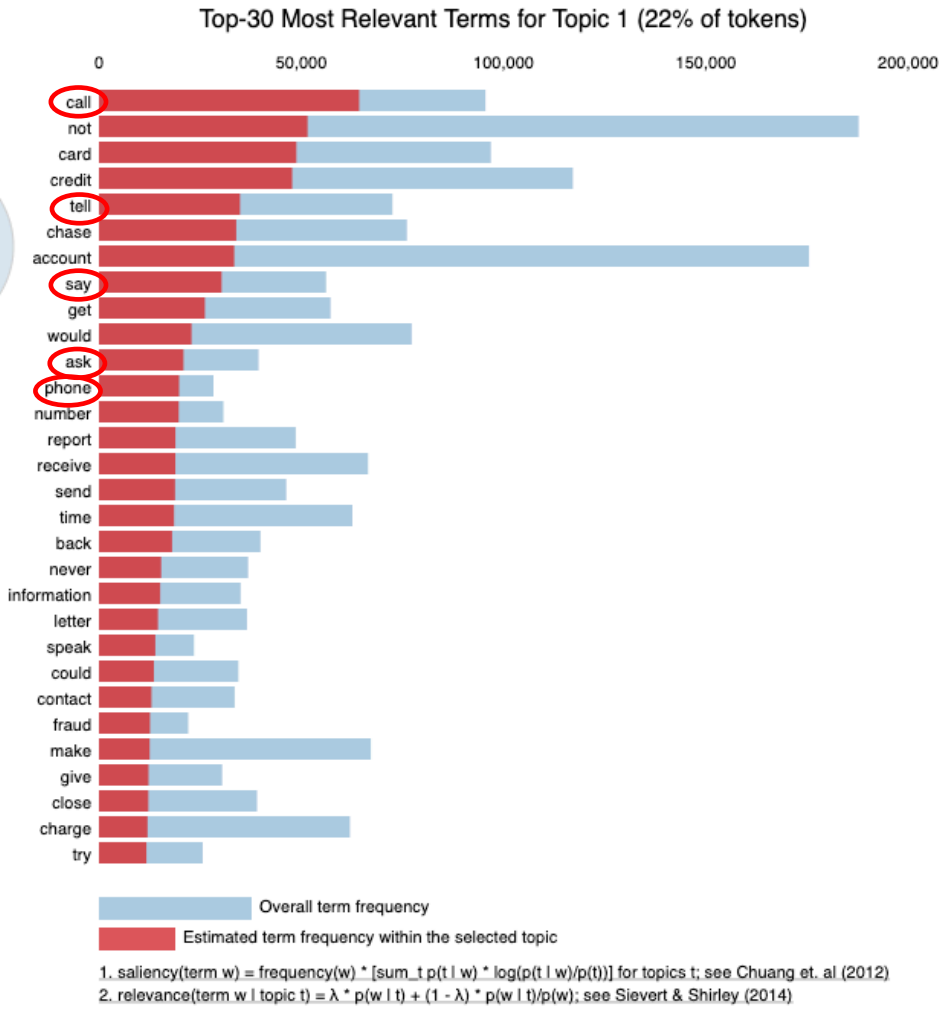






TOPIC MODELING





MODELS USED

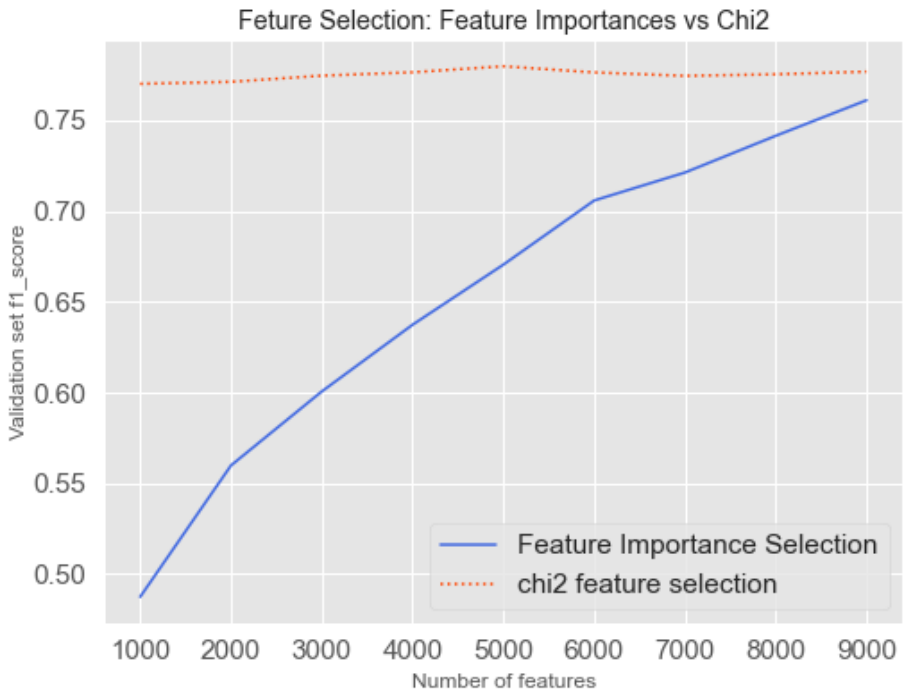
- Random Forest
- Multinomial Naïve Bayes
- Logistic Regression

FEATURE SELECTION

- Chi-squared vs Feature Importances

Random Forest

Naïve Bayes

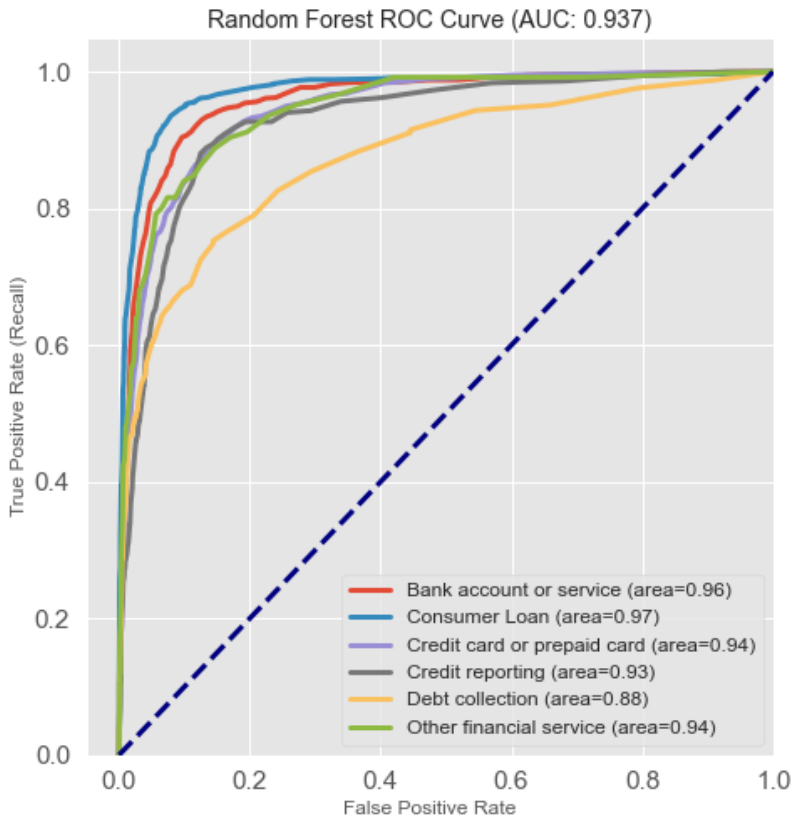


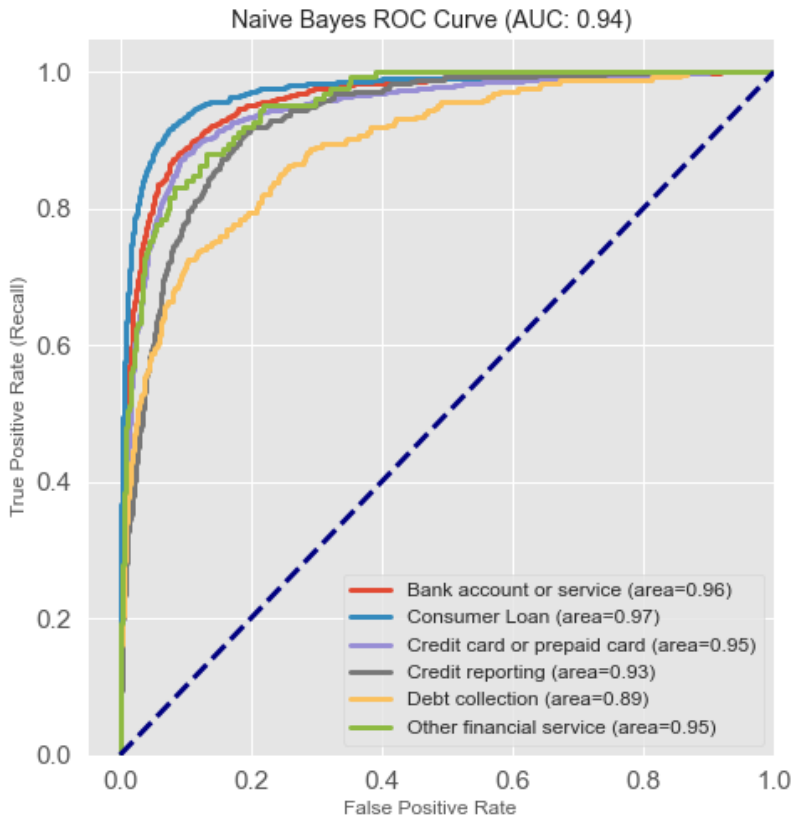
EVALUATION

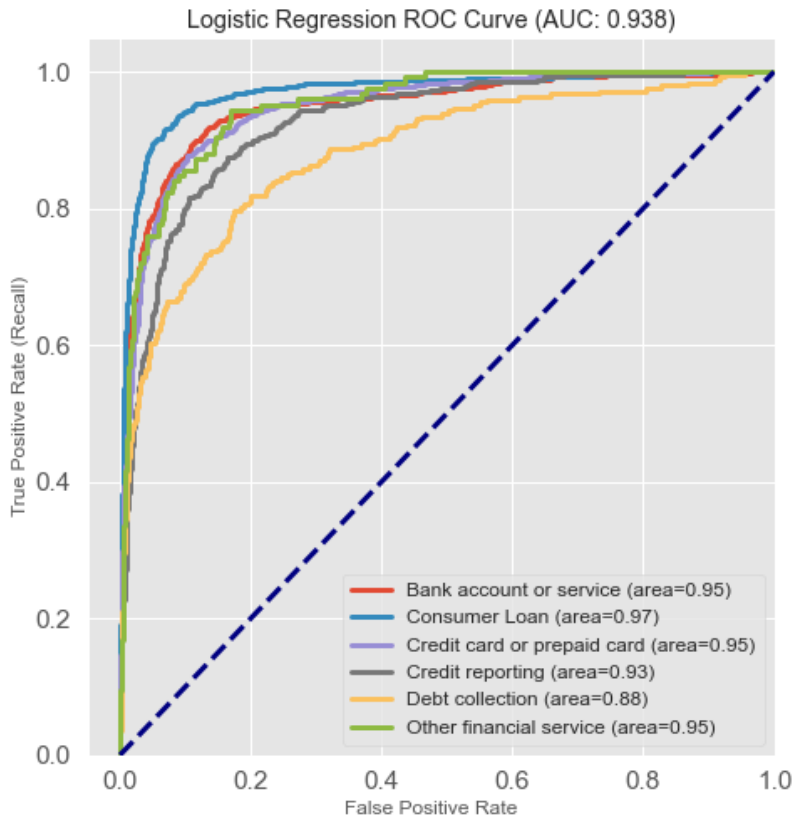
- My recommendation would be to use Naïve Bayes
  - Similar performance to Random Forest and Logistic Regression
  - Very computationally efficient

Model	Weighted f1	Runtime (mins)	AUC Scores
Random Forest Classification	0.787	1.0	0.937
Naive Bayes Classification	0.786	0.0	0.940
Logistic Regression	0.786	22.9	0.938

EVALUATION


Random Forest ROC Curve (AUC: 0.937)
True Positive Rate (Recall)
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Bank account or service (area=0.96)
Consumer Loan (area=0.97)
Credit card or prepaid card (area=0.94)
Credit reporting (area=0.93)
Debt collection (area=0.88)
Other financial service (area=0.94)


Naive Bayes ROC Curve (AUC: 0.94)
True Positive Rate (Recall)
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Bank account or service (area=0.96)
Consumer Loan (area=0.97)
Credit card or prepaid card (area=0.95)
Credit reporting (area=0.93)
Debt collection (area=0.89)
Other financial service (area=0.95)


Logistic Regression ROC Curve (AUC: 0.938)
True Positive Rate (Recall)
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Bank account or service (area=0.95)
Consumer Loan (area=0.97)
Credit card or prepaid card (area=0.95)
Credit reporting (area=0.93)
Debt collection (area=0.88)
Other financial service (area=0.95)

VALUE TO CUSTOMER

- Effectively identify product alignment of complaint
- Compare with the assigned topic to understand a possible other driver

EXPLORE FURTHER

- Neural Network
- Try more techniques to involve Sentiment Analysis

QUESTIONS?