Overview: The Problem

Navigating the loan market can be a treacherous task

How can consumers accurately assess loan companies and the problems they may face when taking out a loan from them?

How can loan companies improve their service based on the complaint that they've received?

Overview: The Solution

Apply natural language processing to data categorized by issue

Allow categorization of past complaints that haven't been categorized

Give lenders and consumers a clearer picture of what problems customers face

Stakeholders: Lenders looking for actionable directions for improvement, consumers who are concerned about specific issues

The Data

Data from the Consumer Financial Protection Bureau

Data includes the text of the complaint, as well as a classification of the issue and the loan company.

Preparing the Data:

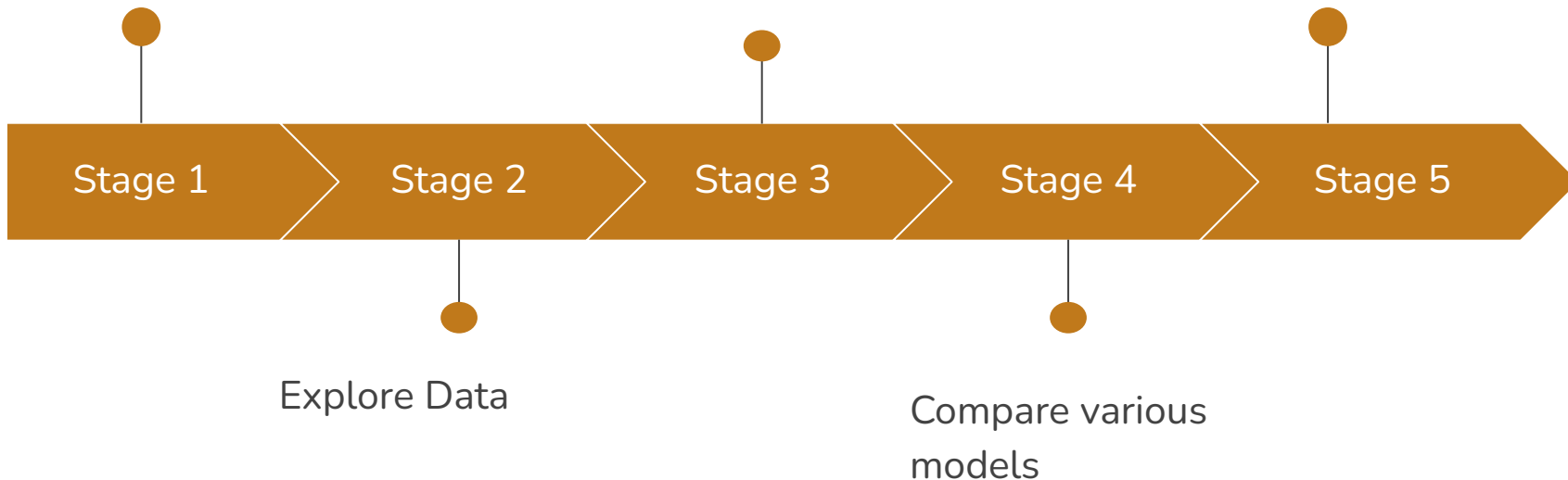- Remove missing values
- Clear duplicate values

The Plan

Clean data, remove
untagged data

Begin fitting various
NLP models to data

Hyperparameter
Tuning

| Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5 |

Explore Data

Compare various
models

Exploring The Data

|	Date received	Sub-product	Issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	Tags	Consumer consent provided?	Company response to consumer	Timely response?	Consumer disputed?
0	2022-03-04	Conventional home mortgage	Struggling to pay mortgage	I can prove that XXXX XXXX lied about my inten...	NaN	Rhode Island Housing and Mortgage Finance Corp...	RI	2904.0	NaN	Consent provided	Closed with explanation	Yes	NaN
1	2022-02-15	Conventional home mortgage	Trouble during payment process	Bank of America did not make the County proper...	Company has responded to the consumer and the ...	BANK OF AMERICA, NATIONAL ASSOCIATION	PA	19365.0	Servicemember	Consent provided	Closed with non-monetary relief	Yes	NaN
2	2022-02-23	Home equity loan or line of credit (HELOC)	Struggling to pay mortgage	I have an equity-line-of-credit with M & T ba...	Company has responded to the consumer and the ...	M&T BANK CORPORATION	MD	20854.0	Older American, Servicemember	Consent provided	Closed with explanation	Yes	NaN

Exploring The Data

Raw Data Missing

Category	Total	Percentage
Tags	2368098	88.3
Consumer disputed?	1913636	71.3
Consumer complaint narrative	1744298	65.0
Company public response	1563346	58.3
Sub-product	235163	8.8
State	39743	1.5
Date received	0	0.0
Product	0	0.0
Issue	0	0.0
Company	0	0.0
Company response to consumer	3	0.0
Timely response?	0	0.0

Mortgage Complaint Issues (Percentage)
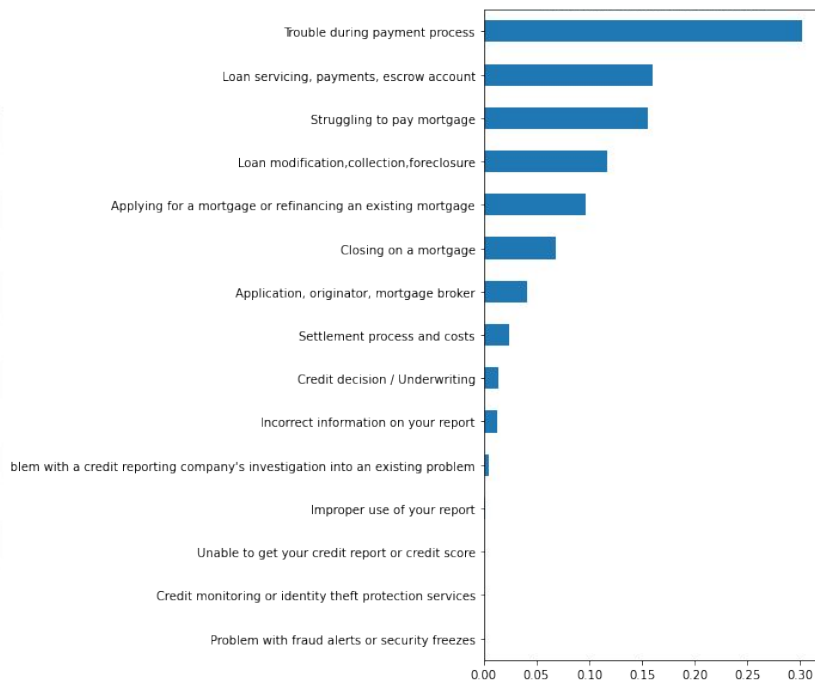
Mortgage Company Responses (Percentage)
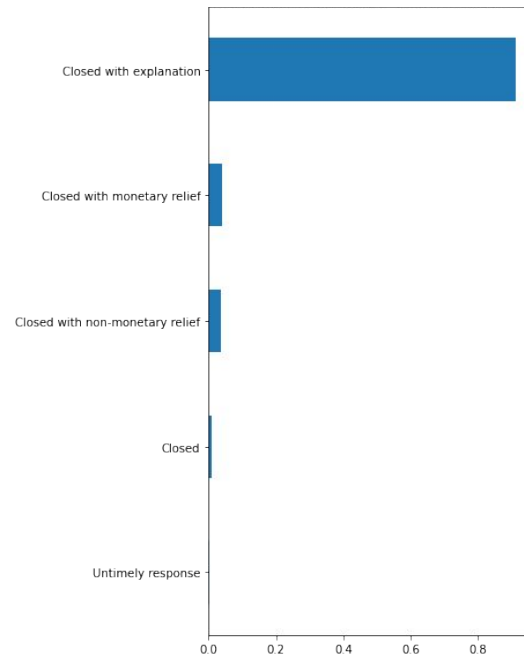
Fitting Initial Models for Supervised Learning

Questions: What percentage of the data should we use to fit our model?

BERT vs RoBERTa vs. XLNET - Which NLP model works best for our data?

Comparing Different Models

After comparison, we arrived at this model

Fitted with 10% of data, using RoBERTa

Micro F1 Score - .69

	precision	recall	f1-score	support
0	0.17	0.12	0.14	8
1	0.60	0.64	0.62	102
2	0.59	0.46	0.52	76
3	0.00	0.00	0.00	4
4	0.00	0.00	0.00	2
5	0.00	0.00	0.00	1
6	0.30	0.62	0.40	13
7	0.36	0.17	0.24	23
8	0.44	0.50	0.47	40
9	1.00	0.12	0.22	8
11	0.00	0.00	0.00	5
12	0.71	0.74	0.73	231
13	0.77	0.80	0.79	403
14	0.00	0.00	0.00	2
accuracy			0.69	918
macro avg	0.35	0.30	0.29	918
weighted avg	0.68	0.69	0.68	918

Directions for Future Work

Finer Hyperparameter tuning

Create dashboard for companies to categorize and explore complaints

Explore data beyond mortgages