

Analysis of Customer Complaints on US Financial Products

Ty Andrews

Dhruvi Nishar

Luke Yang

Table of contents

1	Summary	1
2	Introduction	1
3	Methods	2
3.1	Data	2
3.2	Analysis	3
4	Results & Discussion	4

1 Summary

Here we used multiple classification algorithms to predict whether a financial product consumer will dispute a complaint made to the Consumer Financial Protection Bureaus' (CFPB) Consumer Complaints Database(*Consumer Complaints Database* 2022).

2 Introduction

As of 2022, the CFPB receives over 60,000 consumer complaints a month related to companies financial products. Between December 2011 and November 2022 over 140,000 complaints were disputed by consumers costing both the companies and CFPB time and money.

3 Methods

3.1 Data

Complaints can be responded to by the company in multiple ways but each consumer has the opportunity to dispute the provided response. These disputes are likely costly to both the CFPB and the companies for which they are raised so being able to anticipate whether a complaint will be disputed has the potential to save both time and money for companies and the CFPB alike.



The CFPB database contains ~3 million complaints starting from December 2011 all the way up to November 2022 when this report was written. Approximately 4.8% of all complaints are disputed, with the ~75.1% of complaints having an unknown dispute status. Figure 2 below shows the balance of dispute status.



PLOT PLACEHOLDER IMAGE

Each field has varying amounts of missing values as can be seen in Table 1 below. Fields such as tag where there are numerous entries missing values were removed from the analysis. Individual complaints with missing information were removed from the data-set for analysis since the data set is large enough to still have a significant number of training examples for the analysis (~20,000).

Table 1: Table 1. Unique and missing value counts by data feature.

column 1	column 2	column 3
1	18.30	HELLO
2	17.40	GOOD BYE
3	175.89	AGAIN

3.2 Analysis

A predictive approach using multiple classification models was used to attempt to predict whether a consumer would dispute a complaint or not.

Feature pre-processing approach and rationale is as follows:

Table 2: Feature preprocessing for model training.

Features	Preprocessing Step	Rationale
consumer_complaint_narrative	CountVectorizer max_features = 1000	
product	OneHotEncoder	
sub_product		
issue		
company_public_response		
company		
company_response_to_consumer		
consumer_consent_provided	dropped	

4 Results & Discussion

The results for the different models

Table 3: Table 1. Unique and missing value counts by data feature.

column 1	column 2	column 3
1	18.30	HELLO
2	17.40	GOOD BYE
3	175.89	AGAIN

Table 4: Table 2. Model Performance and Score.

	dummy	logreg	bayes	svc	random.forest
fit_time	1.718	9.357	2.490	170.751	33.682
score_time	0.384	0.703	0.507	34.795	0.865
test_accuracy	0.779	0.655	0.702	0.679	0.783
test_recall	0.000	0.455	0.342	0.445	0.053
test_precision	0.000	0.309	0.332	0.331	0.617
test_f1	0.000	0.368	0.337	0.380	0.097

Consumer Complaints Database. 2022. Consumer Financial Protection Bureau. <https://www.consumerfinance.gov/data-research/consumer-complaints/#download-the-data>.