Exploring differences in Consumer Complaint Behaviour in Financial Products and modelling disputed responses using classification techniques

MSc Research Project
Fintech (Sept 2021)

Ashreet Sangotra
Student ID: x20204523

School of Computing
National College of Ireland

Supervisor: Mr Victor Del Rosal

National College of Ireland
MSc Project Submission Sheet
School of Computing

Student Name: Ashreet Sangotra

Student ID: x20204523

Programme: MSc Fintech **Year:** 2021-22

Module: Research Project

Supervisor: Mr Victor Del Rosal

Submission Due Date: 15.08.22

Project Title: Exploring differences in Consumer Complaint Behaviour in Financial Products and modelling disputed responses using classification techniques

Word Count: 5870 **Page Count:** 26

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:

Date: 15th August 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on the computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Exploring differences in Consumer Complaint Behaviour in Financial Products and modelling disputed responses using classification techniques

Ashreet Sangotra
X20204523

Abstract

Consumers disputing the responses provided by companies to their complaints has been a growing concern, and a pain point for the respective companies. This study aims to build a machine learning model that can predict the likelihood of a complaint response to be disputed. This would help the companies identify those complaints, and proactively take measures to reduce the likelihood of further disputes. The study follows the CRISP-DM approach to this problem. Basic Machine learning classifiers such as Logistic Regression, Support Vector Machine and Random Forest are applied and assessed across various metrics. To improve baseline model performance, sampling techniques as well as NLP techniques have been further implemented. After conducting a thorough analysis across different models and metrics, Random Forest Classifier when applied to randomly under sampled data performs the best, and thus is our best fit model.

Keywords: Consumer complaints, Random Forest Classifier, Logistic Regression, Support Vectors, Natural Language Processing, financial products, consumer behaviour

Table of Contents

1 Introduction

The financial crisis of 2008 was a big turning point for the consumer marketplace, amongst other things. The regulatory environment surrounding financial companies was re-evaluated and re-structured across multiple facets. One such area was consumer protection in the financial sector. The Dodd-Frank Wall Street Reform of 2010 was one of the biggest policy changes undertaken in the financial sector, and sanctioned under this act was the Consumer Financial Protection Bureau (CFPB). Since then the CFPB has been an active agency that monitors how companies in the financial sector use technological tools and mediums to influence and target customers.

Of the many working scopes of CFPB, one of their key offerings is to provide a channel for customers to file a complaint against a certain company's product or service. The customer would file a complaint through CFPB. The Bureau would then keep the company accountable to providing a response to the customer and resolving the issue within a specific timeframe. Once the response has been provided, the customer can further choose to dispute if not satisfied with the response. This has definitely led to higher transparency and accountability in the consumer-company relation in the financial sector. However this also exposes these companies to uncertainty of these disputes and the risks associated with it. These risks could be financial, legal or even reputational in some cases.

The goal of this study is to look for ways that can help companies manage the risk associated with consumer disputes. This is achieved by building a machine learning model that predicts the likelihood of a consumer disputing the response provided by the company.

To achieve this goal, work is done on the Consumer Complaint Database (CFPB, 2021). The Consumer Complaint Database is a collection of complaints about consumer financial products and services that are sent to companies to respond. Complaints are then published after the company confirms and responds to a commercial relationship with the consumer or 15 days later, whichever comes first. Complaints referred to other regulators, such as complaints about depository institutions with less than $10 billion in assets, are not published in the consumer complaints database.

This dataset is a collection of 2.4 million customer complaints in the U.S. that have been filed over a 10 year period - from December 2011 until December 2021. This data is utilised to not only build our classification model, but to also understand some aspects of customer behaviour and larger trends that can be seen over the course of these years. While exploring the data, we are also uniquely positioned to get some insights into the changes in these consumer behaviours and complaint patterns that have occurred in the wake of COVID-19. During our model building stage, the focus will be on a subset of the data, that comprises 768k complaints. This is done so that the complaints that are actually eligible for further dispute can be filtered out. This will also help me make the dataset less skewed, and will need lesser computational resources and time. There are 4 different algorithms that will be built on this dataset. 3 of them are part of the classical ensuite of ML models. They are Logistic Regression, Support Vector

Classifier and Random Forest Classifier. The data pre-processing steps taken for these 3 models are the same. The fourth and final approach is to use language preprocessing techniques (NLP) to build our model. Unlike the previous three models that would use most of the variables present in the dataset, the NLP model would exclusively focus on the actual text complaint written by the consumer.

Looking deeper into the layout of this report, Section 2 aims to review the work surrounding our topic that has already been done in the past. This report will look at studies focused at different aspects of our topic - such as technical learnings for each of these models, studies surrounding consumer behaviour and how it plays out in the financial sector. Section 3 details our methodology, the finer details of our analysis and model building stages, as well the insights gathered as a result of it. Section 4 aims to evaluate the performance of the models across different metrics and techniques. It will discuss these metrics in the context of the business problem that is being solved. In the final Section 5, the study will look at the limitations of our study, the learning outcomes achieved and the scope for further research on this topic.

2 Literature Review

For the last decade or so, consumer complaints have been on the rise(Cormack 2016; Tressler 2016). The ability of the client to express dissatisfaction has never been easier, owing to technologies that can now reduce the burden and social awkwardness of complaining (Dunn and Dahl 2012). Social networking also makes it clear where popular sentiment on individual products and financial services is formed. On the one hand, it may help businesses obtain information and adjust their services accordingly. On the other side, it exposes businesses to public relations and legal dangers(Jung et al 2017). Companies must develop skills to anticipate and respond to consumer displeasure. The Literature review will be divided into 2 sections

2.1 Complaint Behaviour in the Finance Sector Studies

It is critical to have a clear and well-established framework in place for consumers to file complaints. Without one, the company is not only unaware of market demands, but it also results in consumers switching products or propagating negative word of mouth. Criticism assists service providers in resolving issues, satisfying consumers, and improving service quality in the future (Mukherjee et. al 2009).

The competition is fierce, and customers have several alternatives. As a result, customers' expectations have risen (cf. Casado-Díaz & Nicolau-Gonzálbez 2009). The European Commission's Market Monitoring Survey shows that the performance of finance service providers falls short of that of other service markets in general. However, research on Consumer Behaviour Complaints has been rather limited or overly specialised. It has been studied in insurance (Wendel et al. 2011) and banking services (Casado et al. 2011), but these studies were conducted independently. In other situations, the study concentrated on a specific product, such as credit cards (Hogarth et al. 2004), or relied on smaller sets of

data (Ndubisi & Ling 2006). One of our present study's key aims is to examine consumer complaints in the financial industry on a macro level in order to find trends that may be applied to a broader variety of services. Over a 10-year period, we will work with a huge sample of consumer complaints (2.3 million) distributed among 13 commodities and 50 sub-products.

According to Duffy et al. (2006), 41% of respondents reported issues with banking services in the preceding year. One discovery highlighted by Chater et al. (2010) in his study is that consumers devote very little time and effort in learning the specifics of the financial service. Only 48% of investors were familiar with the financial product's features before making a purchase.

Individuals who are young, have a good education, belong to an elite demographic class, have a large salary, and are more socially involved are much more likely to complain since they are more capable, have more self-assurance, and have a higher motive to complain when they are dissatisfied (Tronvoll 2007b).

In his key study from 1988, Singh defined Consumer Complaint Behaviour as "a network of many (behavioural and non-behavioral) behaviour patterns, a few or all of which are prompted by perceived displeasure with a purchase transaction." (Singh 1988, p.94) However, in the financial services market, this strategy falls short since a consumer may be dissatisfied at any moment during the service's existence. Furthermore, with the increasing use of technology and quantitative modelling, taking proactive actions to boost customer pleasure has practically become necessary. The study will develop a classifier that can predict the likelihood of a consumer contesting the company's response to a complaint.

2.2 Consumer Behaviour Studies Using Modelling Techniques

Because of the scope of the information, analysis in the subject has focused on certain subsets of the wider domain. Fosenka et al. (2016) conducted research on the data warehousing aspect of coping with huge files. In the research, the Microsoft SQL server was used for data warehousing and analytics. Naive Bayes, Decision Trees, Neural Networks, and Time Series data mining techniques were used in the study. The quantity of complaint activity in certain financial categories, as well as how the number of complaints fluctuated depending on the economic and political backdrop, was one of the study's key results..

(Silke et al., 2016) investigated the usage of Random Forest for modelling information with ordinal feedback. The study offered two approaches for calculating variable importance that differ from the existing Random Forest methodology for recognizing the value of characteristics. In most instances, normal regression trees performed equally to and somewhat better than classification models in terms of prediction accuracy. This information can be applied to our research.

In their 2016 study, Manisa et al. (2016) performed an in-depth examination and built a model to examine and evaluate the effects of customer dissatisfaction on the operation of the U. S. airline industry (domestic). However, this study limits predictors of company performance to only customer complaints

and ignores issues such as market rivalry, unanticipated occurrences (such as COVID - 19), and so on, all of which had a significant influence in the aviation sector's recent decline.

Unsupervised learning approaches such as Hybrid Clustering are examples of modelling techniques used on customer complaints (Chugani et al., 2018). Xu et al. (2018) use the K Means cluster analysis technique to categorise consumer complaints from mobile providers in another investigation. Investigation of the development of justifications for complaint orders and the subsequent drafting of such orders.

Bastani et al. (2019) used the CFPB dataset to perform Latent Dirichlet allocation (LDA). They analysed the complaint narratives to find latent subjects, which they then examined as time passed. Bastani makes a compelling argument for using LDA to analyse customer complaint narratives by noting four important benefits of the method. Dimensionality reduction, Semantic Annotation, Mixture modelling, and Generalisation are the model's capabilities.

Moedjiono et al. (2016) created a framework for predicting whether or not a consumer would be loyal to a specific brand or product. The study included methods including k-means as well as the C4.5 classification model to get a 79.63% accuracy and also an AUC score of 0.831.

To summarise, each study underlines the importance of understanding customer behaviour across many industries. Though several publications have concentrated on different aspects of customer grievances in the financial sector, few have covered the issue in its entirety and none have studied the events following COVID-19. This is where our research varies from previous studies and fits into a bigger body of work in the subject. To begin, the study will observe shifts in consumer behaviour after the month of March 2020, in addition to how they connect with patterns prior to March 2020. Second, the study will develop a categorization model to evaluate the likelihood of a consumer contesting a company's complaint response.

3 Research Methodology

For our study, it will closely follow the CRISP-DM framework. It is by far the most popular framework used for data mining and Data Science projects. It stands for Cross Industry Standard Process for Data Mining, and is divided into 6 key stages:

The report will start off with gaining more insights about the business problem at hand, and thus define our business objectives. In Data Understanding it will explore the dataset in depth, and focus on answering some questions around the nature of the problem being tackled. The study will visualise and understand larger patterns amongst consumer complaint behaviour in finance. With Data preparation, data will be cleaned and processed so that it’s suitable for building a model. The cleaned data will then be passed onto the modelling stage where it will be trained on 4 different models. After the model is trained, the study will evaluate its performance across key metrics by running the model on test data. This will be done in the evaluation stage. As the focus of this study is model performance, the study won’t delve into the Deployment stage.

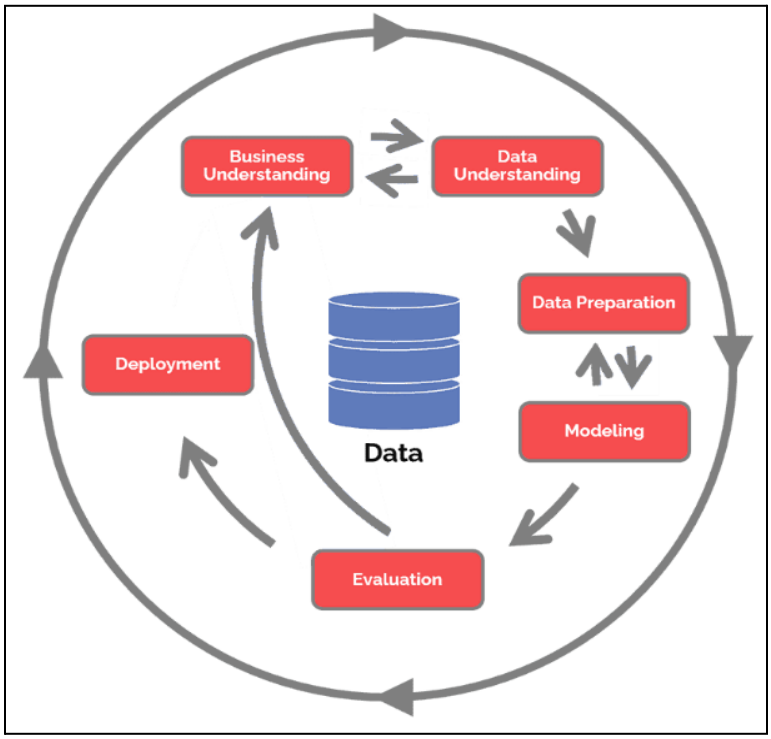


Figure 1: Phases of CRISP-DM Process

(Note: Even though evaluation is a key state in CRISP-DM, it will be discussed in a different section after the Implementation. This is to keep the content in line with the needed report structure)

4 Implementation

4.1 Business Understanding

The number of consumer financial complaints has increased over the past few decades. Our dataset shows that from 2011 and 2021, it grew at an average rate of 29%. Growing complaints do provide the business feedback on their services, but they also put them at danger of litigation and settlements if the client disputes their solutions. Therefore, the ability of the corporation to accurately forecast which complaints are more likely to be contested is crucial. This may enable them to take preventative action on them and in the best interest of both parties involved. Additionally, it would make it possible for the company to distribute resources more effectively. This expresses our business objectives, which can be further defined as being "to decrease the amount of disputed responses." Our data analysis and modelling objectives, which are described in the following sections, may be further derived now that the study has this specific and quantifiable objective.

4.2 Data Understanding

In this section exploratory data analysis (EDA) was performed on the dataset. The visualisations and the insights generated are detailed in this section. However, given the scope of this study, EDA has been limited to just 5 questions

4.2.1 How has the complaint volume changed in the face of COVID - 19?

To answer this question, let's start with plotting a heat map to visualise the complaints received per month over the last decade.

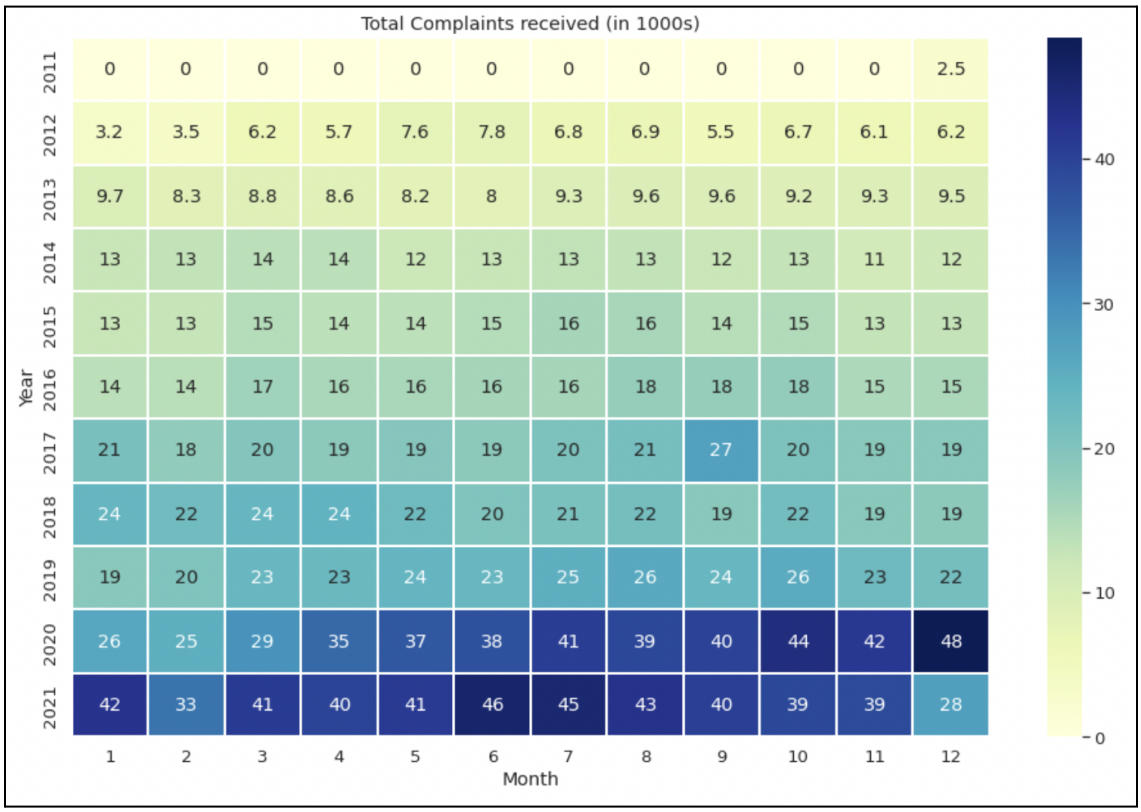


Figure 2: Heatmap of complaint received over different months and years

As it can be seen, the volume of complaints has grown steadily over the years. However after March 2020, the rate of complaints have spiked quite drastically. It's no coincidence that this perfectly aligns with the first wave of COVID. The drastic jump in the number of complaints can be further visualised by a line plot, as can be seen below.

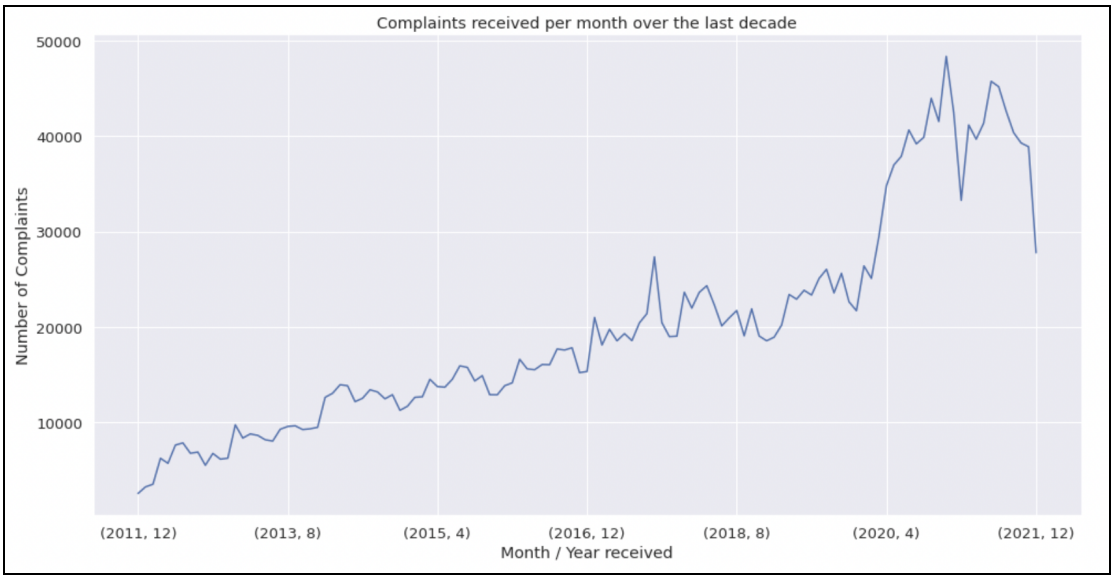


Figure 3: Complaints received per month over the last decade

Another interesting observation that can be made by looking at the heatmap, is the cyclic nature of the complaints. On an average, it seems like more complaints get filed during the summer time, and the general trend of this cycle continues through the years. Coming back to our main question of finding how complaint volume has changed after COVID, I have plotted a density distribution of the complaints filed before 31st March (pre-covid) vs the complaints file after 31st March (post-covid).

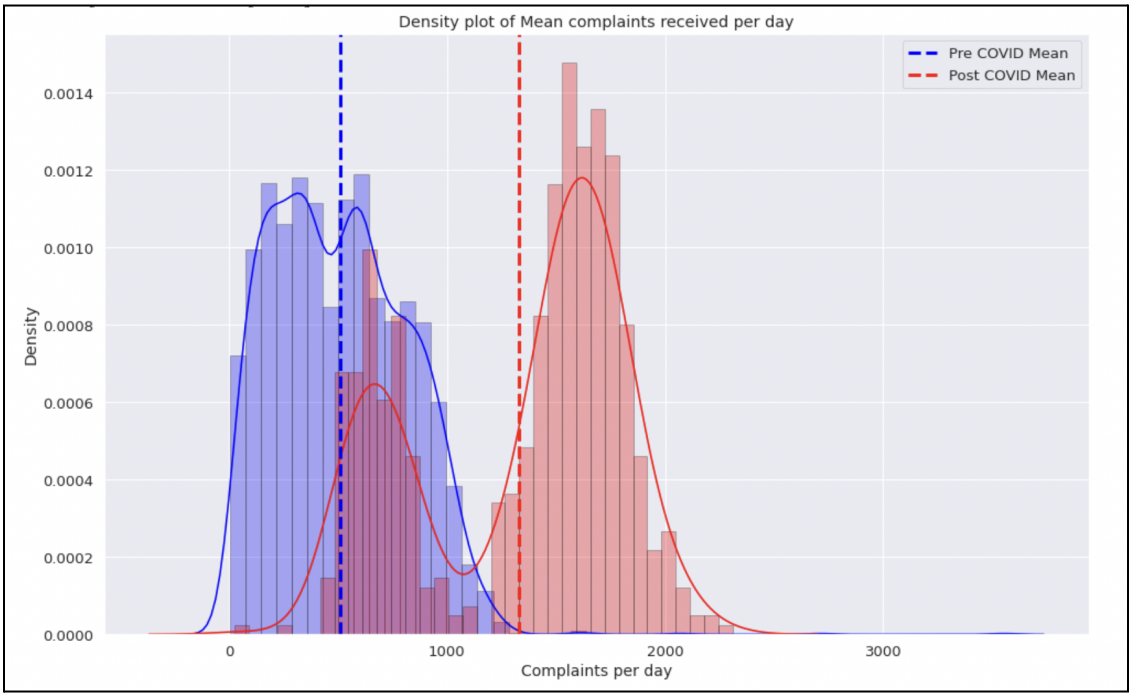


Figure 4: Density plot of Mean complaints received per day

Mean complaints received per day before 31st March: **510.89**
Mean complaints received per day after 31st March: **1331.3**

The mean complaints received per day have more than doubled.

4.2.2 How are the complaints geographically distributed across the country?

Given the wide spread and distribution of population across the country, this is generally reflected in the geographical pattern of complaints that can be observed in the dataset.

As seen below, California is the state with the highest number of complaints, followed by Florida and then Texas. The top 5 states account for over 40% of the total complaints lodged.

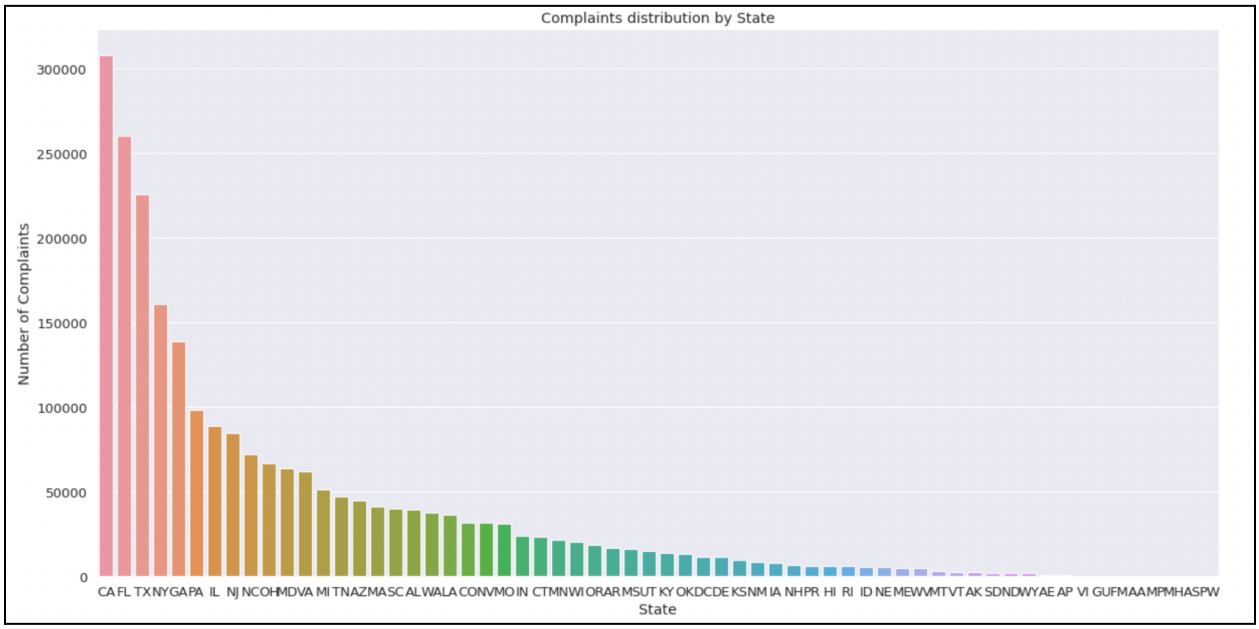


Figure 5: Complaint distribution by State

The author has further plotted the distribution of complaints across the map of the U.S to give a better representation of the spread. Both west and east costs seem to have a more denser complaint pattern as compared to the mid regions of the country.

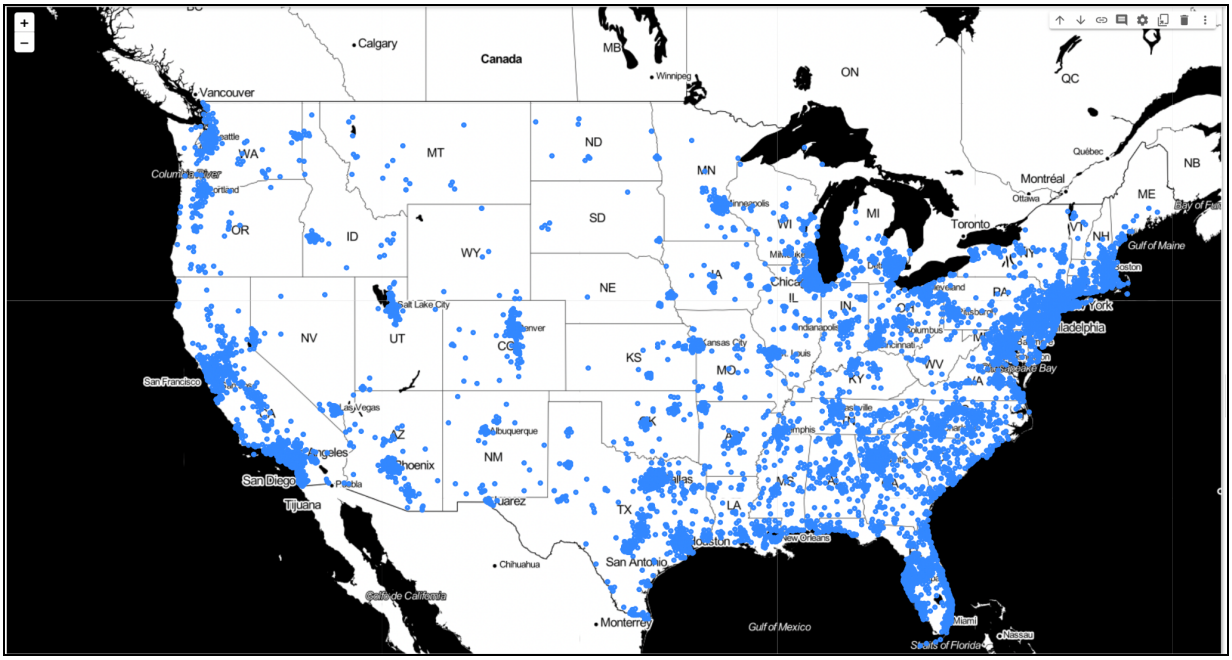


Figure 6: Complaint distribution across the country

(Note: Due to limited computational resources, the above graph has been plotted using 1% data sampled from each state)

4.2.3 Which products have the most complaints and disputes against them?

Certain products and services account for a much bigger chunk of the total complaints filed as compared to others. In the graph below it can be seen that the percentage contribution of each product category. Mortgages have the highest percentage of complaints attributed to them. In fact the top 3 products contribute to more than 50% of the total complaints. These are Mortgage, Credit Reporting and Debt Collection. The proportion of complaints disputed for each product can be difficult to assess from this plot, and is thus addressed in the next visualisation.

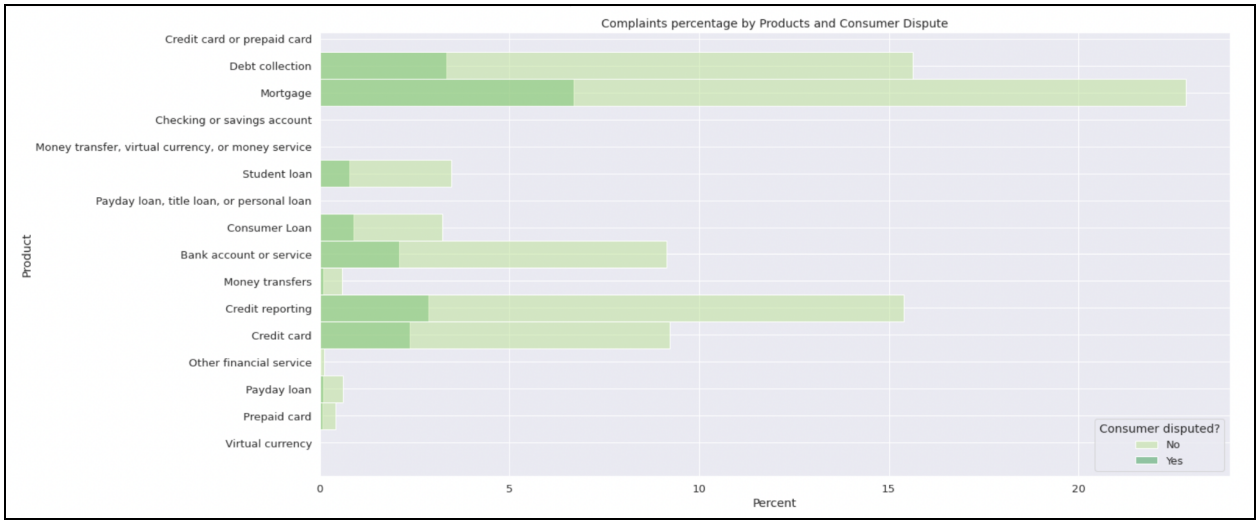


Figure 7: Percentage of complaints by products and consumer disputes

Over here, there is a plot of the percentage of complaints that were disputed within each product category. Though the mean disputes centre around 20% of total complaints, mortgages can be seen to have a higher than average proportion of disputed complaints. However it is Virtual currencies that have more around 50% of total complaints being disputed. This could be attributed to the decentralised nature of the blockchain technologies, the fraudulent activities reported on it, and the lack of regulation around it.

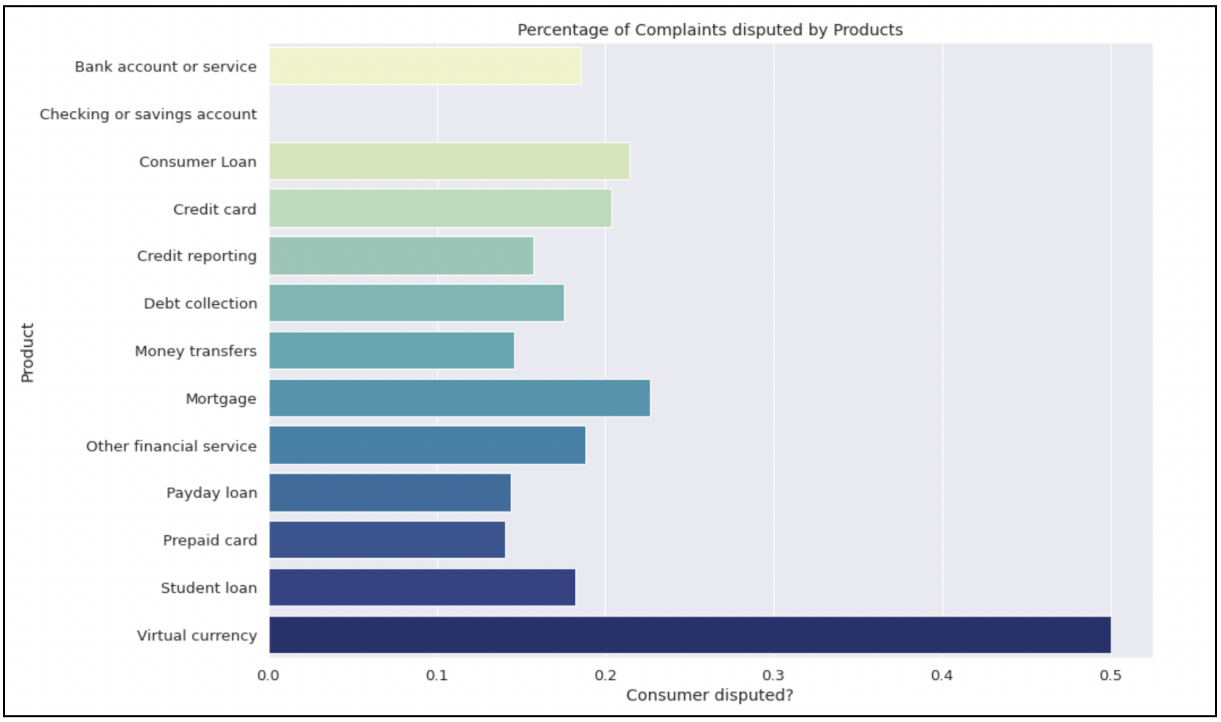


Figure 8: Percentage of disputes by product

4.2.4 In the wake of COVID-19, are more complaints being lodged online?

Given the massive shift of communication channels to digital mediums in the wake of the pandemic, It makes intuitive sense that after the same shift would have been observed for consumer complaints as well. This assumption is easily validated as the bar plot showing the percentage distribution of companies before and after COVID. Now, more than 85% of the complaints are lodged through the web. On the other side of the spectrum, it can be seen that Postal mail has had the most significant drop.

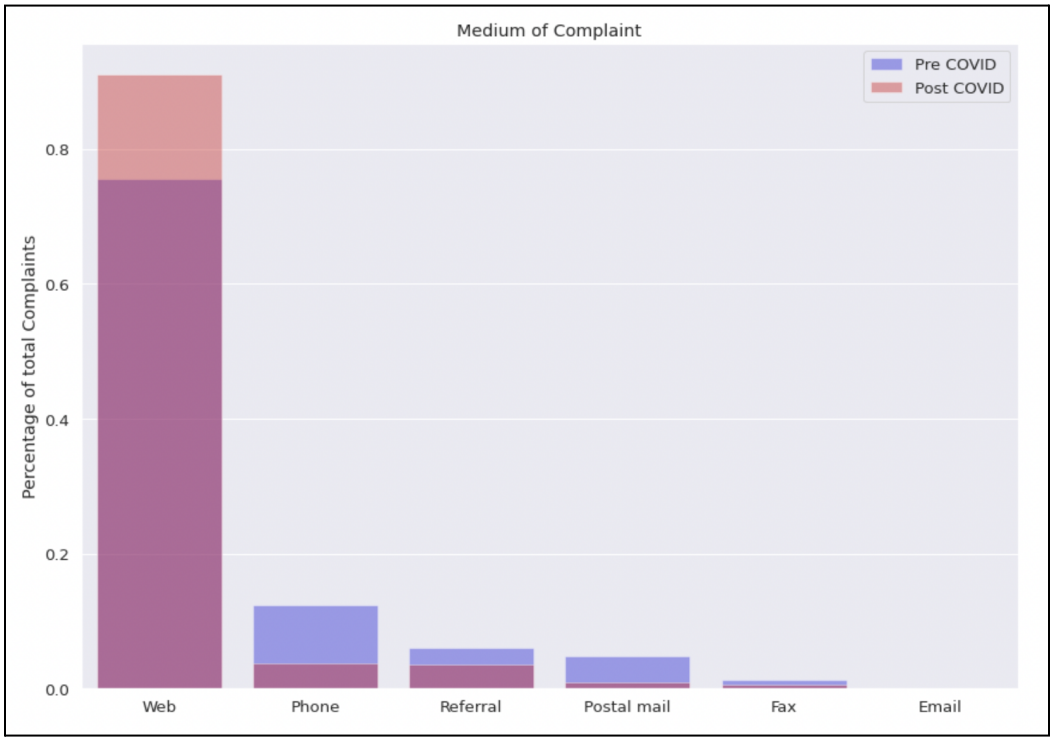


Figure 9: Medium for lodging Complaints

4.2.5 Which products tend to have the most elaborate complaints?

One of the features of our dataset is the text complaint written by the consumer. In the bar plot below, the average length of the complaint for each product is shown. The length represents the total characters in the complaint string. As it can be seen, complaints related to mortgages are by far the most elaborate. Not only are mortgage based complaints more elaborate, as seen earlier - Mortgage is also the product with the highest number of complaints against it.

Of the top 5 products, 3 of them are related to huge loan repayments (Mortgage, Vehicle Loan and Student Loan). It can be observed that the bigger the financial commitment, the more concerned and detailed consumers are with their complaints. Whereas, complaints regarding short term borrowing (payday loan) or reporting take the bottom of the list.

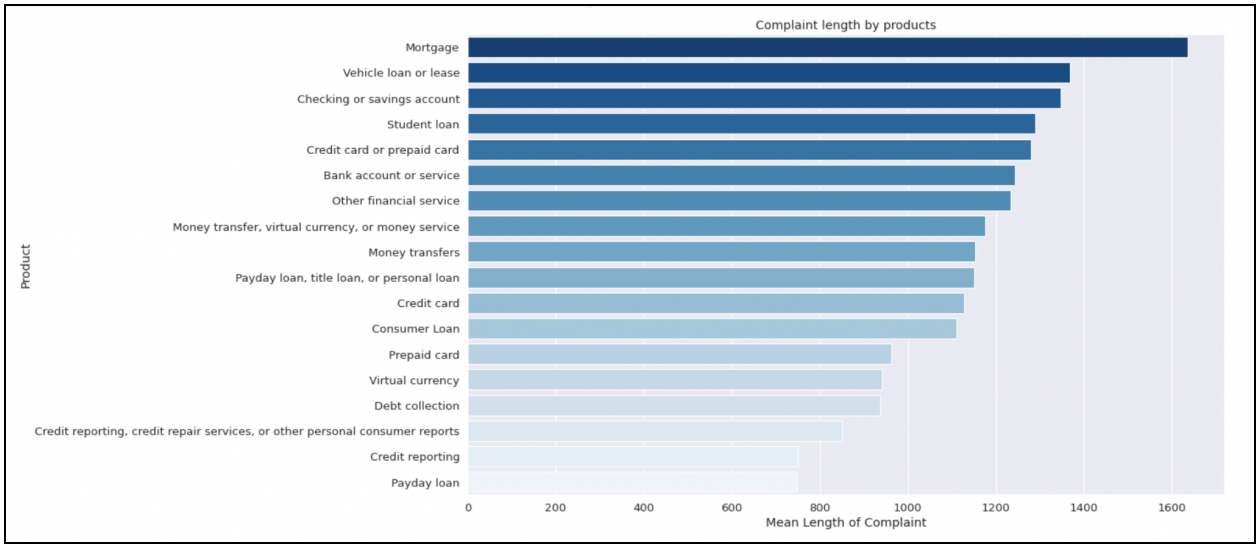


Figure 10: Mean length of complaints by product

4.3 Data Preparation

In our dataset, 768k of the 2.4 million rows have information for consumer disputes. While the entire dataset was used to perform EDA, focus will be on the 768k complaints while building our model. In this study, 2 different methods for data preprocessing have been adopted. They are:

1. Traditional Data pre-processing
2. Natural Language Processing

This will allow us not only to compare performance across different models, but also across different processing techniques. This also ensures the utilisation of as much relevant information as can be extracted from the dataset. While the report starts by looking at all the features for traditionally processing the data, it shall only focus on ‘Consumer Complaint Narrative’ for the NLP method.

The data processed through these techniques will each be modelled on 3 different classifiers and evaluated across 3 different metrics. An overview of the same can be seen below.

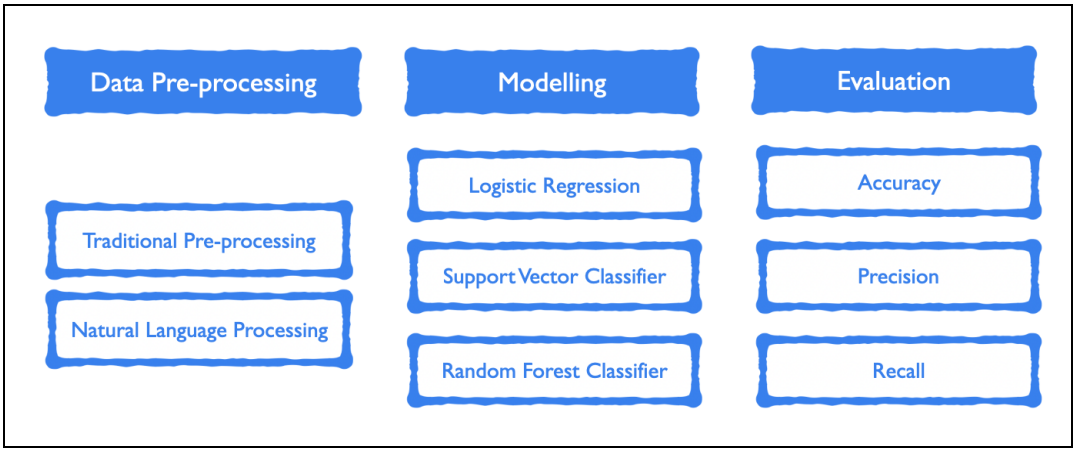


Figure 11: Overview of Model Building

4.3.1 Traditional Data Processing

For traditional pre-processing and cleaning of the data, each of the features will be looked at individually. The table below details the steps undertaken for treating each feature. After cleaning the data and applying One Hot Encoding (OHE) to categorical variables, our pre-processed training set comprises 308 columns in total. A breakdown of that can be seen in the table as well.

Feature Name	Missing Values %	Unique Values	Used for Model Building	Pre-processing needed	No. of columns in training set	% of total columns
Date received	0.00%	1970	Yes	Split into month and year features. Then One Hot Encoding (OHE)	19	6.2%
Product	0.00%	13	No	Drop Column. All info here of this feature is encapsulated within sub-product	0	0.0%
Sub-product	30.60%	50	Yes	All Missing values correspond to 3 unique products, implying those products don't have sub products. Replace NaN values by product name itself.	50	16.2%
Issue	0.00%	99	Yes	One hot encoding (OHE)	99	32.1%
Sub-issue	59.26%	61	No	Drop Column. Too many missing values	0	0.0%
Consumer Complaint Narrative	78.65%	160968	No	Drop Column. Too many missing as well as unique values.	0	0.0%
Company Public Response	74.32%	10	Yes	Replace missing values with 'No public response'. Then OHE	11	3.6%
Company	0.00%	4285	Yes	Except for the 50 most mentioned companies, replace all other companies with 'Other'. Then OHE	51	16.6%
State	0.74%	62	Yes	Drop Rows with missing values	62	20.1%
Zip Code	0.73%	44857	No	Drop Column (too many unique values)	0	0.0%
Tags	85.88%	3	No	Drop Column (too many missing values)	0	0.0%
Consumer consent provided?	61.23%	4	No	Drop Column (insufficient info)	0	0.0%

Submitted via	0.00%	6	Yes	One hot encoding (OHE)	6	1.9%
Date send to company	0.00%	2049	Yes	Make new feature 'Days to process' by calculating the difference between this feature and 'Date received'	1	0.3%
Company response to consumer	0.00%	7	Yes	One hot encoding (OHE)	7	2.3%
Timely response?	0.00%	2	Yes	Label encoder	1	0.3%
Consumer disputed?	0.00%	2	Yes	Target variable. Replace 'Yes' and 'No' with 1 and 0.	1	0.3%
Complaint ID	0.00%	768459	No	Drop Column. No valuable information here.	0	0.0%

Table 1: Feature Engineering (Feature overview and actions taken)

4.3.2 Natural Language Processing

As opposed to traditional Data pre-processing, the focus will only be on the actual complaint text here. The feature which captures this is 'Consumer Complaint Narrative'. The following steps were taken to correctly structure and clean this text data:

1. **Label Encoding:** The 2 labels in the target variable are encoded into numbers. 'Disputed' being represented by 1, and 'undisputed' being represented by 0.
2. **Tokenizing Text:** Extract all the unique words used across all the complaints narratives, and assign them a number (aka a token).
3. **Text to Sequence:** Each complaint is now represented as a sequence of tokens. These tokens correspond to the words they were assigned to in the previous step.
4. **Sequence Padding:** All sequences of all complaints are padded as rows. The resulting array is then used as the features of the dataset.

4.4 Modelling

Now that the data has been pre-processed and cleaned, the next step is to build our ML models on it. Model building is a computationally expensive task. Building a model on 768k rows would be practically impossible on a local machine. For this reason, data is subsetted before going ahead. As the size of datasets are different at the end of the 2 pre-processing approaches, 2% of the dataset of the traditional pre-processed dataset is used, and 5% for the dataset processed for Natural Language. Do note these are 2 different datasets and the models will be trained on each of them separately.

The first step to do here is to split the data into a train and test set. The conventional 80 / 20 split between train and test sets is followed. The second thing to note here is the imbalanced nature of the dataset. Only 21% of the complaints in our dataset are disputed. So if the model would just predict all complaints as 'undisputed', it would be right 79% of the time, and have an accuracy of that. This implies a false narrative about the performance of the model. To allow the model to learn more accurately, sampling techniques were applied. Given larger size of data is computationally challenging on a local machine, sampling was focused towards under-sampling techniques instead of over-sampling. Three

under-sampling techniques - Random Under Sampling, Cluster Centroids and Tomek Links have been applied on the training set. The data sampled from each of these techniques would then have 3 different models built on it. After training different models across different preprocessing and sampling techniques, the report can now look at comparing model performance across all of them. This has been done in the Evaluation section.

4.5 Hardware and Software Specifications

The hardware that will be used for this project is a MacBook Air M1 (2020). Hardware specifications of the machine are mentioned below:

Model: MacBook Air M1, 2020, 256 GB
Processor: Apple M1
RAM: 8 GB
Graphics: Apple M1 GPU Built-In 7-core

The project will explicitly use Python for conducting all parts of the project - Data cleaning, querying, visualisations and modelling.The implementation of the entire project will be done in Python and hence the software requirements will be a Jupyter Notebook with Python version 3.7.2. Main Python Libraries that will be used are Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn and Statsmodel.

5 Project Evaluation

5.1 Overview

Now that the models have been trained, the next step is to evaluate their performance. There are 3 metrics the study will use for this - Accuracy, Precision, Recall.

While accuracy simply measures the percentage of total correct predictions and can be easily manipulated in the case of imbalanced data, Precision and Recall allow us to take a more detailed view into the performance of the model.
Precision answers the question - Of all the positive predictions made, how many were actually true.
Whereas recall answers - Of all the actually true cases, how many were predicted as true.

There is always a trade-off between Precision and Recall. Trying to improve one negatively affects the other. Let’s have a look at the performance of our different models across these 3 metrics. Following which the study will discuss which metric should be focused on given our business problem, and thus which is the best bit model amongst the lot.

Traditional Data Pre-processing					
Sampling Technique	Model	Time Taken	Accuracy	Precision	Recall
Random Undersampling	Logistic Regression	0.27	55.11%	55.00%	23.08%
	Support Vector	23.13	57.01%	59.17%	24.96%
	Random Forest	0.72	52.20%	64.17%	23.63%
Cluster Centroids	Logistic Regression	0.45	51.54%	66.17%	23.73%
	Support Vector	20.42	45.09%	70.00%	21.92%
	Random Forest	0.85	26.54%	95.17%	20.51%
Tomek Links	Logistic Regression	0.74	80.21%	0.83%	35.71%
	Support Vector	129.85	80.34%	0.00%	0.00%
	Random Forest	2.22	80.34%	0.00%	0.00%

Table 2: Performance of models under traditional data processing

Natural Language Processing (NLP)					
Sampling Technique	Model	Time Taken (secs)	Accuracy	Precision	Recall
Random Undersampling	Logistic Regression	0.32	52.41%	48.48%	22.86%
	Support Vector	16.76	56.19%	55.37%	26.52%
	Random Forest	1.89	54.97%	56.47%	26.08%
Cluster Centroids	Logistic Regression	0.28	57.28%	35.54%	21.64%
	Support Vector	12.39	34.86%	70.52%	21.02%
	Random Forest	2.34	37.36%	59.23%	19.63%
Tomek Links	Logistic Regression	0.76	75.44%	10.74%	33.05%
	Support Vector	91.14	77.88%	0.28%	50.00%
	Random Forest	5.41	77.76%	0.00%	0.00%

Table 3: Performance of models under Natural language processing

From the above 2 tables, sampling done by Tomek Links has the best accuracy. But it scores 0 on Precision and Recall. What this would mean is that this model is predicting everything as 'undisputed'. And since undisputed complaints compromise around 80% of the data, the model is getting the predictions right 80% of the time. However this is misleading as the model has no predicting power over the 'disputed' complaints.

5.2 Precision vs Recall in the context of the Business Problem:

Let's look at the first row in the Tradition Data Preprocessing table as an example. The recall score is 23.08%, meaning the model only predicts 23.08% of the 'disputed' complaints as 'disputed'. However of all these 23.08% of complaints it does predict as 'disputed', 55.0% of them were actually 'disputed'. The remaining 45% were False positives.

With higher recall, more of the actual 'disputed' complaints can be captured, but in this process the model will mark a lot of 'undisputed' complaints as 'disputed'. This would redundantly increase the workload and resource allocation on the side of the companies as they would have to deal with a lot of False Positives. However with a higher precision, the model will not capture all 'disputed' complaints, but the ones it does predict as 'disputed' would have a huge probability of being true. Though this is still the perfect scenario for the companies, it does allow them to potentially reduce the disputes against them by around 20%. The Study will choose precision as our metric of choice. Fortunately our models in the tables above naturally seem to lean towards higher Precision.

Going back to our tables, Random Forest trained on traditionally preprocessed data and using Cluster Centroids as the sampling mechanism has the highest Precision (95.17%), which is quite good. This will be selected as our best fit model.

5.3 Fine tuning for Precision

ML models predict the complaints as 'disputed' or 'undisputed' by calculating the probability of the complaint to be 'disputed'. By default the probability threshold is at 0.5. This means if the probability that a complaint will be disputed is greater than 0.5, then it will be predicted as 'disputed', and as 'undisputed' if the probability is lesser than 0.5.

Changing this threshold can in turn change the model performance. The graph below shows us how precision, recall and accuracy changes for our best fit model as there is change in the probability threshold.

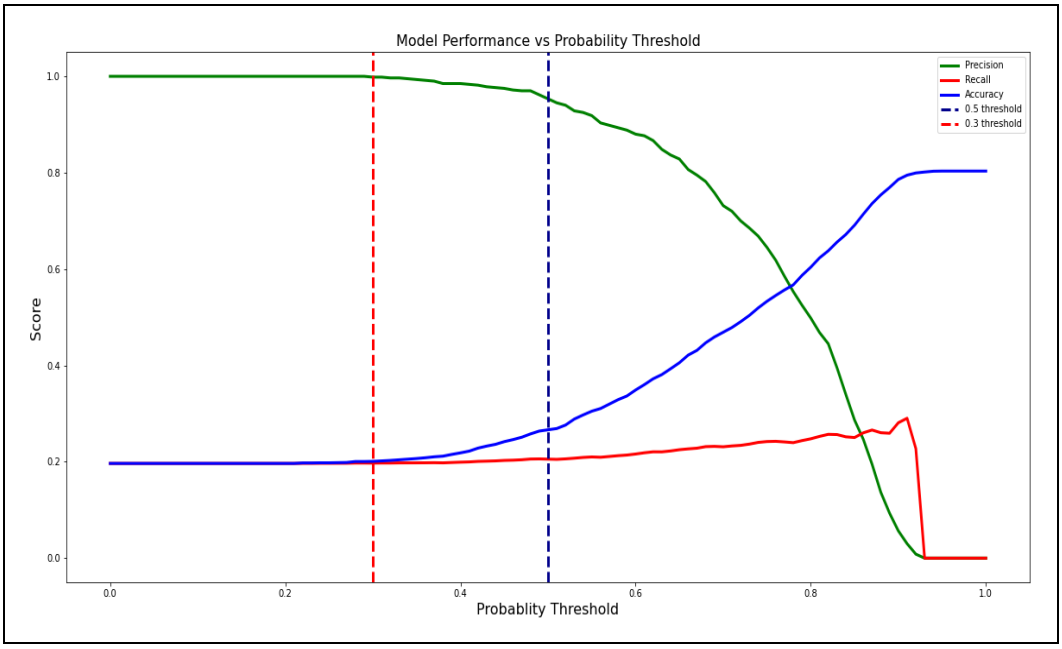
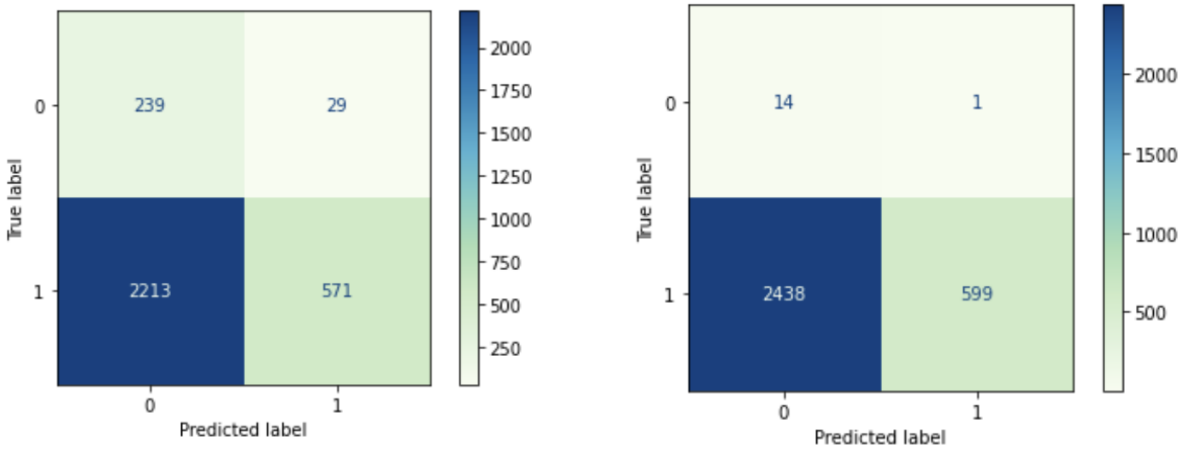


Figure 12: Best-fit model performance across different thresholds

While Precision drops as the threshold goes from 0 to 1, the recall pretty much stays in a small range for most of the probabilities. Instead of going with the default threshold of 0.5 (indicated by the blue dashed line in the graph), the threshold will be finteuned to 0.3 (shown by the red dashed line). At this threshold recall pretty much stays the same, however precision almost approaches 1 (99.8%).

A comparison of confusion matrices before and after threshold fine tuning for the best fit model is shown below. A significant reduction in False Positives from 29 to 1 can be seen..



(Before Fine tuning) **(After Fine tuning)**

Figure 13: Confusion Matrices before and after threshold tuning

5.4 Discussion

Our study starts off by exploring different aspects of the data to gain some insight into consumer complaint patterns. After covering this, the study delved into modelling the disputed responses by consumers by use of various classification techniques. Though the study made significant headway into improving our performance metrics, there is still a long way to go before the solution could resemble anything close to a commercial product. The limitations of our study are mentioned below:

1. **Deployment:** The final stage of CRISP-DM is to deploy the model to a specific use-case. However this would have extended the scope of the project multifold and would require development skills.
2. **Scale of Data:** As all the modelling was done on a local machine with limited processing power, the size of the data used for training was extremely downsized.
3. **Sampling constraints:** Our study exclusively stuck to under-sampling techniques due to the limited volume of data that could be processed. Oversampling techniques such as SMOTE are usually mode accurate
4. **Use of ANNs:** When it comes to massive datasets with tonnes of features, Artificial Neural Networks are the best suited algorithms for these. However they need lots of computing power, and it could take upto weeks to run multiple epochs on our complete dataset.
5. **Low Recall and Accuracy:** One of the biggest shortcomings is the low recall and accuracy of our best fit model. Despite being precise, its poor performance over other metrics really limits the scope of using it in a commercial environment.

A second interaction of effort into this topic could help to work on the above mentioned limitations. Earlier in our study, Exploratory Data Analysis was performed to understand different aspects of our data. It was seen how the complaints received per day spiked to more than twice after COVID-19 hit. It was also seen how the east and west coast of the U.S had higher complaint density. California, Florida and Texas were the biggest 3 states by complaint volume. Mortgage was the product with the most number of complaints, whereas virtual currencies had the highest number of disputes. Unsurprisingly, COVID-19 increased the complaints lodged over the web, and decreased the ones lodged through post. Finally te report saw that consumers tend to be the most elaborate while lodging complaints around Mortgage, student loans and vehicle loans. Higher the financial stake, the more concerned consumers are.

This research may be utilised by companies who offer financial services and even by the fintech community, whose goal is to give customers greater financial service. This will assist both parties in making an educated decision on customer service to prevent problems with their goods or services in the future. Additionally, based on the present circumstances surrounding the topic under consideration, this research may be utilised to assess the likelihood or potential that the corresponding customers may contest the respondents' claims.

6 Conclusion and Future Work

This study had an objective of modelling disputed complaints in the financial sector. With extensive modelling done across 2 data preprocessing techniques, 3 undersampling techniques and 3 classification models; the study was able to compare performance between 18 unique trained models.

Random Forest Classifier when trained on traditionally preprocessed data that has been undersampled using Cluster Centroids was our best fit model.

After getting our best fit model, it was optimised further by fine tuning the probability threshold for predictions.

Among the outcomes is that consumers' behaviours are not significantly changed by prompt answers from financial service providers to their complaints about financial goods and services. This demonstrates that customer happiness may be considered to be a determinant of the quality of financial goods and services. In other words, customer complaints are a tool to gauge the quality of financial services. This is seen in how different nations, notably the United States, manage problems involving financial goods in relation to customers' happiness.

As mentioned in the last section, our study had quite a few limitations. This in turn leaves a good scope for future work to be conducted on this topic. Further work should be done in using oversampling techniques on the complete dataset to improve model performance. This would require industrial grade computing and the use of GPUs. Advanced NLP techniques such as Stemming and Lemmatization could be used to preprocess the data more thoroughly. Finally Neural Networks such as CNNs (Convolution Neural Networks) might prove to be effective algorithms in training a dataset of so many features and categories.

Acknowledgement

I would like to extend my sincere gratitude to Mr Victor Del Rosal, whose consistent and extremely well structured supervision went a long way in helping me carry out this study. The regular touchpoints not only made sure I was on schedule with the study, but also gave crucial insight to proceed ahead whenever stuck.

I would also like to thank Mr Noel Cosgrave who served as my supervisor in the 2nd Semester while I was deciding on my research project. His guidance allowed me to pick a relevant topic that I successfully carried into the Research Semester.

7 References

Bastani, K., Namavari, H. and Shaffer, J. (2019). Latent dirichlet allocation(lda) for modelling of the cfpb consumer complaints., Journal of Expert System with Application 27: 256–271.

Casado-Díaz, A.B. and Nicolau-Gonzálbez, J.L. (2009), "Explaining consumer complaining behaviour in double deviation scenarios: the banking services," *The Service Industries Journal*, 29 (12), 1659-1668.

CFPB (2021) *Consumer Complaint Database.* Available at https://www.consumerfinance.gov/data-research/consumer-complaints/#download-the-data (Accessed: 11th March 2022)

Chater, N.Huck, S. and Inderst, R. (2010), *Consumer Decision-Making in Retail Investment Services: A Behavioural Economics Perspective*, European Commission.

Chugani S, Govinda K, Ramasubbareddy S. "Data Analysis of Consumer Complaints in Banking Industry using Hybrid Clustering". In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC) 2018 Feb 15 (pp. 74-78)

Cormack, Lucy (2016), "Consumer Complaints Increase by Almost 10 Per Cent in a Year," Sydney Morning Herald, August 15, http://www.smh.com.au/business/consumer-af fairs/consumer-complaints-increase-by-almost-10-per-cent- in-a-year-20160815-gqsq8q.html.

Duffy, J.A.M., Miller, J.M. and Bexley, J.B. (2006), "Banking customers' varied reactions to service recovery strategies," *International Journal of Bank Marketing*, 24 (2), 112-132.

Dunn, Lea and Darren W. Dahl (2012), "Self-Threat and Product Failure: How Internal Attributions of Blame Affect Consumer Complaining Behavior," Journal of Marketing Research, 49 (5), 670–81.

European Commission (2012a), *Consumer Markets Scoreboard. Making Markets Work for Consumers. 8th edition – December 2012*, European Union, Brussels.

Fonseka, W. R., Nadeesha, D. G., Thakshila, P. M., Jeewandara, D. M. and Wijesinghe (2016). 'Use of data warehousing to analyse customer complaint data of consumer fin- ancial protection bureau of united states of america', in 2016 IEEE International Con- ference on Information and Automation for Sustainability (ICIAFS), Galle, Sri lanka, 16-19 dec 2016, pp. 152-158, IEEE Xplore. doi: 10.1109/ICIAFS.2016.7946520.

Jung, K., Garbarino, E., Briley, D. and Wynhausen, J. (2017). Blue and red voices: Effects of political ideology on consumers complaining and disputing behaviour, Journal of Consumer Research 44(3): 477–499.

Manisa, S. U., Anitsal, M. and Anitsal, I. A. (2016). A model of business performance in the US airline industry: how customers' complaints predict the performance, Journal of Business Studies 8(2): 96–111.

Moedjiono, S., Isak, Y. and Kusdaryono (2016). 'Customer loyalty prediction using k-means segmentation and c4.5 algorithm, in 2016 International Conference on In- formatics and Computing(ICIC) , Mataram, Indonesia, 28-29 oct 2016, pp. 210-215. IEEE Xplore. doi:10.1109/IAC.2016.7905717.

Mukherjee, A., Pinto, M. B. and Malhotra, N. (2009), “Power perceptions and modes of complaining in higher education,” *The Service Industries Journal*, 29 (11), 1615-1633.

Silke, J., Gerhard, T. and Boulesteix, A. (2016). Random forest for ordinal responses: Prediction and variable selection, Journal of Computational Statistics and Data Ana- lytic 96(3): 57–73.

Singh, J. (1988), “Consumer complaint intentions and behaviour: Definitional and taxonomical issues,” *Journal of Marketing,* 52 (1), 93-107.

Tressler, Colleen (2016), “Consumer Complaints to the FTC Increased in 2015,” Federal Trade Commission, March 1, https://www.consumer.ftc.gov/blog/consumer-complaints- ftc-increased-2015.

Tronvoll, B. (2007b), “Complainer characteristics when exit is closed,” *International Journal of Service Industry Management,* 18 (1), 25-51.

Xu J, Zhao J, Zhao N, Xue C, Fan L, Qi Z, Wei Q. “The Research and Construction of Complaint Orders Classification Corpus in Mobile Customer Service”. In CCF International Conference on Natural Language Processing and Chinese Computing 2018 Aug 26 (pp. 351-361). Springer, Cham.