Machine Learning models for Customer Relationship Analysis to Improve Satisfaction Rate in Banking

Patricia Fabijanczyk   Lahari Bagam   Nan Jia

Abstract—The purpose of this research project was to analyze customer complaint data from financial institutions and identify areas of opportunity for these institutions to improve their customer satisfaction rate. In addition to pointing out areas for improvement, this paper also looks into similar research and tries to understand if themes found in this analysis are consistent with those done by other researchers. Banking is an essential piece to everyday life for all people across the world. Banks need to ensure that their products and processes are simple and accessible to all. Although banks have a monopoly on our financial needs their desire to retain existing customers and gain new ones drives the necessity of providing excellent and timely customer service.

The study was conducted using a dataset of over two million customer complaint records and examining what were the top three financial institutions receiving complaints and which products received them. In addition, other aspects of complaints such as state of origination was also looked at.Analysis was done using machine learning, python, tableau and other tools to show the data points and their correlation. Understanding the top financial institutions methods of handling customer complaints, we are able to make recommendations for further product improvements to increase customer satisfaction. Concluding the research project is a list of challenges and opportunities for further research projects. In addition, there are recommendations for the financial institutions investigated in this project on how to move forward from analyzing customer complaint data.

Index Terms—Machine learning, Finance, NLP, online information

1. Introduction

With the development and increased use of technology providing feedback has become instant and easy. Feedback can be negative or positive but to companies it should be what they constantly ask for and want to receive. Customer feedback can also make or break a company [1]. With either positive or negative customer feedback a company can capitalize on what customers offer to them via multiple channels and incorporate it back into their innovation chain. Time and time again we have seen companies that do not listen to their customers fail – whether it be they continue to

exist but struggle to gain momentum or they fail completely and are forced to shut their doors. One prime example of this is the sad fate of Blockbuster; although they had plenty of opportunities to reimagine the way they brought movies to families across the United States they took the stance that they knew their customers and they would stick to what they knew – brick and mortar stores with VHS tapes [1]. Netflix entered the scene and completely changed the landscape of at-home entertainment and eliminated their first competition, Blockbuster [1]. Although it is difficult to compare services between Netflix and Blockbuster to financial institutions, fundamentally they serve to provide a service to us the customers, whether it be entertainment or keeping our money secure in a checking account or helping us finance a home [1]. The Blockbuster example is a prime example of why companies should ask for and utilize customer feedback. Financial institutions are a staple to our lives but with banks competing to differentiate themselves there are institutions that can fold to others by not evolving to meet customer needs.

1.1. Research Paper's Structure

This research paper is structured to explain the importance of customer satisfaction to a financial institution. Section two focuses on explaining the problem in detail and its significance to the market. Section three highlights the dataset used to conduct the analysis and how it was prepared for processing. Section four discusses the approach and methods used to analyze the large dataset. Section five, and six dive deeper into published research found over the course of the project that guided and informed this research. Section seven investigates the data analysis and tools used to conduct the research, such as using machine learning to develop a series of graphs from the data set. Section ten rounds off the research project by highlighting some challenges in the project and opportunities to expand the research while provide recommendations to the top three banks on where focus should be for improvements within their product lines to ensure customer satisfaction remains high and drives higher customer retention in return.

Banking is a necessity everyone must interact with on a regular basis. The days of cash only transactions are over and financial institutions Capital One, Bank of America, Wells Fargo, BB&T just to name a few, now take on the

role of intermediary when it comes to us and our assets. Each of the financial institutions given as an example and many more provide us with a number of services when it comes to finances – checking and savings accounts, loans, mortgages, credit cards. If we do not like something in the services, we receive we naturally turn to the company we are experiencing problems with and voice our dislikes in hopes that it will alter the way we do business with them moving forward and avoid similar issues in the future. Complaining to a financial institution can be done through multiple channels – via the web, over the phone, even sometimes over their app or on social media.

The market today offers numerous companies to pick from when it comes to banking – there are small town banks, credit unions, and large financial institutions that offer a variety of products to customers. When customers have a number of options companies strive to differentiate themselves in the market and win customers – either by providing competitive rates for loans, checking rates with a higher than average interest rate, rates on mortgages, incentives for opening up credit cards, and quality customer service [2]. Customer satisfaction counts for a lot today than all of the products a company can provide [2]. "If your customer is not satisfied, he or she will stop doing business with you [2]." Customer satisfaction is defined as the customer's "perception that his or her expectations have been met or surpassed [2]." For banking, customers expect their funds to be available to them when needed, that includes making sure the technology that supports customers is working properly. When customers have a good experience with a company with a product it is more likely that they will return the next time they have a need for that product or a new one [2]. Financial institutions want customers to build portfolios with them, it results in greater profits for them, so customer satisfaction is what every bank strives for.

Since interacting with financial institutions is such a fundamental part of everyone's lives it is important to do further research and analysis into what customers struggle with to identify areas of opportunity for banks – not only for their product development but also for overall customer service and interaction. There is plentiful literature available on the importance of customer satisfaction and the impact on business if it is low. Each financial institution might define success has something different, but it is at the heart of a bank to serve a customer for all their financial needs so in our analysis and research we will look to provide insight into areas that they can focus on improving to hopefully reach new customers or encourage existing customers with other products to expand their services.

2. Dataset

The first dataset was sourced from DataWorld [3]. The dataset consists of fifteen thousand records. Information that is provided includes: date complaint was received, product and sub-product, issue with product and complaint narrative, company response, company name, state and zip code of complaint origination, complaint submission

method, time response, and whether the consumer disputed or not. The data fields that will be used for analysis will beproductand sub-product, issue with the product, company response, company name, state of complaint origination, complaint submission method, and consumer disputed field. All records that are not complete with these fields will be excluded from evaluation. Complaint narratives will be used for a more detailed description of the issue submitted. The dataset spans the years 2011 to 2019 allowing us to do a year-over-year comparison of complaint volumes in certain products and states.

The second dataset was sourced from the Consumer Financial Protection Bureau and has over two million records to analyze for this project [4]. Fields from the first dataset can be mapped to the fields in the second dataset thus creating a larger sample to run analysis on. Things to note about the CFPB dataset are that complaints in this set are only published after the company responds and the relationship with the consumer is confirmed [4]. Data points collected and that appear in both data sets are: company name, company's response to the consumer, whether the response was timely (a yes or no response), date and state received, the product the complaint is against, and the issue, some complaints have additional details provided in the form of sub-issue [4]. A full list of data provided can be seen in Table 1.

Not all datapoints in the dataset were used during this research project. The most important datapoints and those that were used in the analysis were: company, product, sub-product, issue, sub-issue, complaint ID, and state. Timely response was considered as an additional aspect for correlation but responses were only limited to 'yes' and 'no' which do not provide enough to draw out any conclusions. Additional details for timely response would have required criteria that would define what timely is for each bank.

2.1. Data Cleaning, Processing and Input Tools

We have processed our data in order to find if there are any duplicate or null values present in our dataset. While processing the data, we found that there were no duplicate values but there were two null values that were removed in order to visualize the data in Tableau.

For initial data analysis we used Tableau (Table 2 for initial code) to create some visualizations for a subset of the data set. Moving forward we plan to make use of Python Jupyter Notebook and some of their libraries like Seaborn and Pandas for achieving the results using machine learning techniques. For Python, we will use Pandas and Matplotlib. For R, we will use tidyverse and ggplot2.

3. Approach and Method of Ecaluation

With our current dataset we will be able to extract the top customer complaints by banking institution and the method they were received. The data set also provides the state the customer was from, the date the complaint was

TABLE 1. COMPLETE LIST OF COMPLAINT DATA CONSUMER FINANCIAL PROTECTION BUREAU PUBLISHES. [1: DIRECT PULL OF DESCRIPTIONS FROM CFPB]

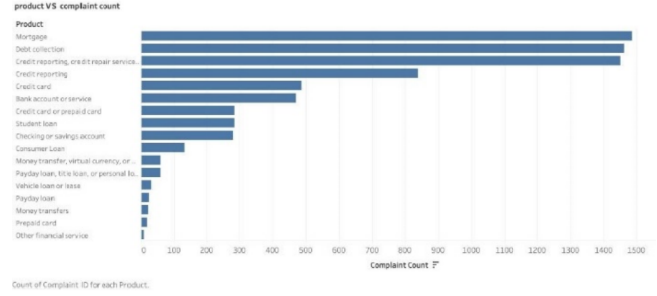| Field Name	Description
Date received	The date the CFPB received the complaint. For example, "05/25/2013."
Product	The type of product the consumer identified in the complaint. For example, "Checking or savings account" or "Student loan."
Sub-product	The type of sub-product the consumer identified in the complaint. For example, "Checking account" or "Private student loan."
Issue	The issue the consumer identified in the complaint. For example, "Managing an account" or "Struggling to repay your loan."
Sub-issue	The sub-issue the consumer identified in the complaint. For example, "Deposits and withdrawals" or "Problem lowering your monthly payments."
Consumer complaint narrative	Consumer complaint narrative is the consumer-submitted description of "what happened" from the complaint. Consumers must opt-in to share their narratives. We will not publish the narrative unless the consumer consents, and consumers can opt out at any time. The CFPB takes reasonable steps to scrub personal information from each complaint that could be used to identify the consumer.
Company public response	The company's optional, public-facing response to a consumer's complaint. Companies can choose to select a response from a pre-set list of options that will be posted on the public database. For example, "Company believes complaint is the result of an isolated error."
Company	The complaint is about this company. For example, "ABC Bank."
State	The state of the mailing address provided by the consumer.
ZIP code	The mailing ZIP code provided by the consumer. This field may: i) include the first five digits of a ZIP code; ii) include the first three digits of a ZIP code (if the consumer consented to publication of their complaint narrative); or iii) be blank (if ZIP codes have been submitted with non-numeric values, if there are less than 20,000 people in a given ZIP code, or if the complaint has an address outside of the United States).
Tags	Data that supports easier searching and sorting of complaints submitted by or on behalf of consumers. For example, complaints where the submitter reports the age of the consumer as 62 years or older are tagged "Older American." Complaints submitted by or on behalf of a service member or the spouse or dependent of a servicemember is tagged "Servicemember." Servicemember includes anyone who is active duty, National Guard, or Reservist, as well as anyone who previously served and is a veteran or retiree.
Consumer consent provided?	Identifies whether the consumer opted in to publish their complaint narrative. We do not publish the narrative unless the consumer consents, and consumers can opt-out at any time.
Submitted via	How the complaint was submitted to the CFPB. For example, "Web" or "Phone."
Date sent to company	The date the CFPB sent the complaint to the company.
Company response to consumer	This is how the company responded. For example, "Closed with explanation."
Timely response?	Whether the company gave a timely response. For example, "Yes" or "No."
Consumer disputed?	Whether the consumer disputed the company's response.
Complaint ID	The unique identification number for a complaint.

Figure 1. Count of Complaint ID for each Product

received, status of the complaint and whether the complaint was addressed in a timely manner.

Text analysis for investment and sentiment analysis of online comments can be use for improving user satisfaction in online banking [5], [6], [7], [8], [9], [10] . In this work, we impliment bussiness investment and customer satisfaction by both machine learning and NLP models based on the NLP models developed by our research mentors for textual online finance information [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23] . We will be using data visualization to show the breakdown of customer complaints by financial institution, the type of product and number of complaints it receives, and the number of complaints received by state. Using data visualization, we will be able to: tell which product received the most complaints, identify the financial institutions with the most complaints, look across the financial institution and products and determine which state produces the most complaints.

Also, we use applications of machine learning by natural language processing for finance and bussiness as well [18], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39] in other domains such as health, security and business, and apply transfer learning models to analyse the online comments for creating a model for emergency response [?], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53]. From this data building out a decision tree will help us look at what financial products are offered through the financial institutions and map out potential customer complaints that could be received based on the dataset being analyzed.

Initial data visualizations of a sample size of seven thousand records show that Mortgages received the greatest number of complaints closely followed by Data Collection and Credit Reporting (Figure 1). Adding an additional layer to complaint analysis we are able to see that the sub-product that receives the most complaints is credit reporting (Figure 2). This is reflective of the data that was analyzed by the Consumer Financial Protection Bureau (Figure 5) which also captured credit reporting as the product that received the most complaints. In addition, we see in Figure 6 that the number of complaints captured at the national level also mirror our initial results that complaints originate mostly in the Eastern part of the United States.

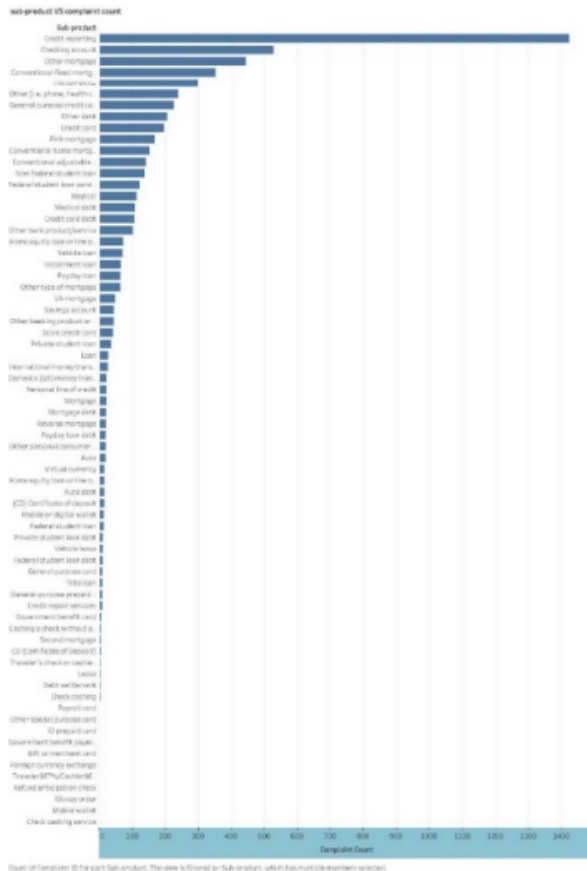Looking at the sample dataset we are able to see that

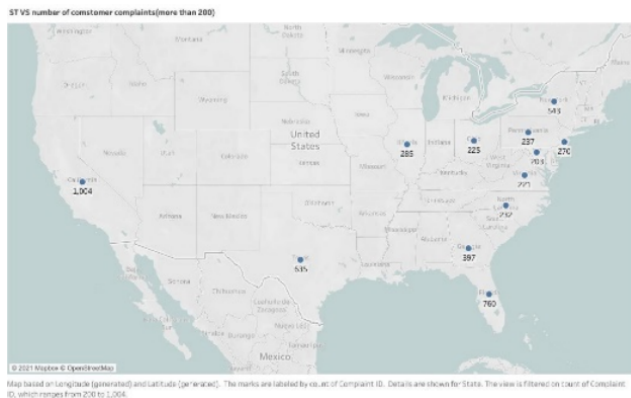Figure 2. Count of Complaint ID for each Sub-Product

Figure 3. States that received ¿ 200 complaints

complaints are mostly concentrated in the Eastern states of the U.S. with a significant number also originating from California and Texas (Figure 3). Complaints breakdown for states was based on if they received more than two-hundred complaints. Tableau was used in order to process a sample population from our dataset and create the data visualizations. Tableau will continue to be a main tool used in order to help show the data in an effective and digestible
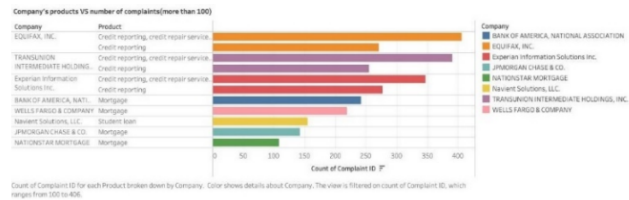

Figure 4. Company's Products vs the Number of Complaints ¿ 100

TABLE 2. CODE USED FOR DATA CLEANING IN TABLEAU

```
# import needed libraries
import pandas as pd # for data frame creation
import os # for OS interface (to get/change directory)
# display and set working/data directory
os.getcwd()
os.chdir('\Volgenau School_Data Analytics\Fall2021\AIT-582\project')
os.getcwd()
# import the data;
data = pd.read_excel('Modified_dataset.xlsx')
# data information and check duplicates and null values
data.info()
data.head(10)
data.duplicated()
data.isna()
```

manner.

In addition, based on this dataset we will be able to classify consumer complaints, which are in text data, into the following categories: debt collection, consumer loans, mortgage, credit card, credit reporting, student loans, bank account or service, payday loan, money transfers, prepaid card, and other services. These categories will assist in identifying the areas of issues and will help with predicting product and issues using the historical complaint data. From this, we will be able to provide plotting graphs, scatter plots, histograms, and correlation matrix.

4. Reason for big data solution

A big data solution will allow for greater detection about customer sentiment regarding a company's products or services. Using Machine Learning will allow for a more robust way to process information [54]. The machine power can process the dataset in order to calculate various types of variables from the population [54] at a faster and more efficient way. In addition to overall sentiment, we will be able to look at multiple facets of the complaints received, such as, channel complaint was received through, the company that received the complaint, the state from which the complaint originated from, and whether or not it was resolved in a timely manner. Similarly, the Consumer Financial Protection Bureau uses the large dataset of complaints to analyze it and help govern and guide companies with more informed financial laws, rules and regulations [4].

5. Published Research

To understand how a company can be successful one needs to begin by understanding what complaints most are received. Merriam-Webster defines complaint as "[an] expression of grief, pain, or dissatisfaction". "Customer experience is the emotion felt by customers1 when they come into any contact with a company – no matter how or by what means. It is what customers remember from their interaction with a business. [55]" There are a number of studies that have been done that show that customers that have a positive experience with a company are 86% more likely to return and do business with that company again and goes up to 92% when that customer is already an existing customer [55]. In addition to those higher percentages, losing customers can be very costly for a business. In another study done, 50% of customers said they left because they did not feel valued and had a poor experience with customer service [55]. Another staggering number is that 80% of customers would willingly pay more for a product or service if the customer experience is better [55]. If those numbers are not enough to convince a company the value that customer satisfaction provides then it has been analyzed that acquire new customers can cost five to twenty-five times more than retaining an existing customer [55].

5.1. Customer Satisfaction & Measurement

To effectively measure and predict customer satisfaction one must define how to measure it. An important piece to measuring satisfaction is also understanding quality as it plays a key role in how the customer determines their expectations of the product they are interacting with [2]. Quality is difficult to define in the realm of customer satisfaction for the simple fact that it can vary person-to-person – thus making it more difficult to quantify [2]. One way to gather measurement for customer satisfaction is through the use of surveys [56]. Surveys give companies the insights directly from their customers. "Strong relationships lead to higher levels of loyalty with customers [thus] resulting in profitability [56]". Service quality, when looking at the "Relationship Survey Framework" (Figure 5), depends on the following customers experiences: installation, complaint, billing, purchase, pre-sales, and any other ones they interacted with [56]. When all those experiences are positive then it builds or improves upon the relationship resulting in an overall better product experience, price, and increase in corporate reputation [56].

In another study where the researchers looked at the impact of customer satisfaction and its effect on customer relationship management showed that customer satisfaction increased by nearly 50% with proper CRM [57]. In addition, it was called out like mentioned earlier in this paper that customers are key for the success of a company and the company must "optimize customer's satisfaction and minimize various problems which correlate directly to customers" [57]. This study was done on customers and personnel of freight companies although it is not a stretch
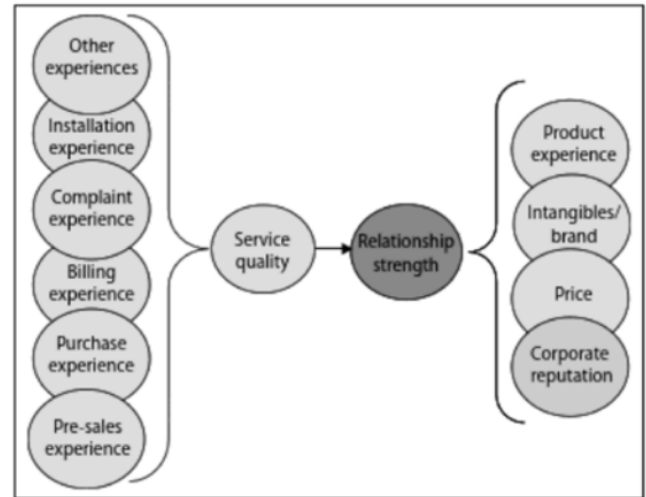

Figure 5. The Relationship Survey Framework [56]

to hypothesize that the positive correlation between CRM and customer satisfaction across other industries, such as that of the financial world, where banks strive to have loyal customers that invest in more than one product from them [57].

5.2. Customer Complaints Leading to Business Downfall

Since covering the importance of customer satisfaction it is also imperative to look at how neglecting this aspect has led to a company's downfall. The demise of Blockbuster was covered in the Introduction of this paper. Although technology was a huge reason for why Blockbuster ceased to exist it was also the lack of understanding of its customers that played a large role [1]. "Blockbuster came to be seen as a giant corporation that charged outrageous sums for late fees and cared little for what its customers wanted the most [1]." Additional examples of businesses that failed due to their lack of attention to customers are W.T. Grant and US Airways [1]. W.T. Grant was a large American retailer during the early years of 1900s eventually coming to an end in the 70s after they failed to adapt to their customers needing more retail space located closer to their neighborhoods – this came about as more families moved out from cities and relocated to the suburbs [1]. US Airways literally slashed their budget when it came to customer service and outsourced this part of the business – this "cost-cutting" tactic only failed to satisfy customers as their complaints and questions were not being addressed [1]. All three of these examples show just how important the relationship with the customer is and if not handled effectively then it could lead to bankruptcy like it did to US Airways, Blockbuster, and W.T. Grant [1].
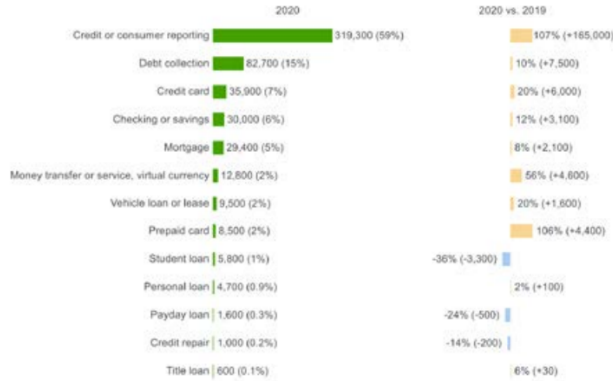
Figure 6. "Complaint Volume by Financial Product or Service" [2]

Figure 7. "U.S. Complaint Submissions per 100,000 Population" [2]

5.3. Government Oversight & Protection

Most might not see complaints as a productive piece of society, but the Consumer Financial Protection Bureau created the "Consumer Complaint Database" which compiles customer complaints received about financial products and/or services [4]. Although a c single complaint cannot be taken as a blanket experience for all users collecting data at the government level allows for greater oversight and helps deduce consumer insights and help "regulate consumer financial products and services under existing federal consumer financial laws, enforce those laws judiciously, and educate and empower consumers to make informed financial decisions. [4]" In addition to creating and maintaining this database the CFPB published the 'Consumer Response's Annual Report to Congress' which provides additional accountability on the companies to respond appropriately to customers.

From the Annual Report published this year for 2020 we are able to see a breakdown of the top complaints for the year; at the top being credit or consumer reporting (Figure 6 for reference, taken from the 2021 Annual Report); which is up by "107% from the previous year". This data will be interesting to compare to our dataset and see if what our records indicate mirror customer sentiment nationally. In addition to the top complaints, the CFPB analyzed complaints received in each state – Figure 7 shows the states that had the most complaints originate from [58]. Another aspect of our analysis will be to measure of the volume of complaints originating from states.

A new aspect of the 2021 Annual Report that was not previously discussed was the impact the pandemic which we will not be analyzing in our data set. In addition to the consumer complaint database maintained by the Consumer Financial Protection Bureau the Office of the Comptroller of the Currency exists to ensure financial institutions stay in compliance of consumer laws and regulations and if an institution fails to comply the OCC takes actions as necessary [59]. The OCC "assist[s] national bank customers with questions and complaints [59]" in addition to providing advisory services to customers to explain their rights [59].

Unlike the CFPB, the OCC does maintain a public accessible consumer complaint database but if needed official records can be requested [59].

5.4. Study on Online Customer Complaints in Turkish Banks

Upon researching, a similar study was done on banks in Turkey that had received complaints via online. Their study states: "It is imperative to analyze this large amount of online customer complaints data and to evaluate the obtained results by the banking sector. Online customer complaints, which are usually transmitted in text, are large volume, scattered and have complex data structure" [60]. Similarly to our project, words from the complaints captured were identified and used together to create the scope and make inferences from the complaints made [60]. Their research also found that banks that had largecustomer portfolios where within the top that received the most complaints, this can be seen in the analysis completed in this project as well [60]. Bank of America, Wells Fargo, and JPMorgan Chase & Co. offer their customers a variety of products resulting in customers with larger product portfolios [61], [62], [63]. When looking at the studies word/attributes that most appeared in complaints they included words such as: branch, blocked, cancellation, cancel card, credit card, interest, etc. [60]. These terms slightly deviated from the commonly seen words in our dataset which can be seen in Figure 8.

5.5. Complaints & Customer Satisfaction Analysis in Indonesian Banking

In addition to the study on Turkish Banks, a similar study was completed on Indonesian Public Banking [64]. "The study presents a proposed conceptual model, which is a key determinant of customer loyalty. The results show that the quality of the service has a positive effect on satisfaction, but the quality of service does not affect customer loyalty" [64].

An interesting aspect of this research study on Indonesian banks is the analysis of genders and age ranges submitting complaints [64]. In addition to those two customer attributes, education level and career information was also gathered and analyzed [64]. The results of the study showed that the quality of service had a positive impact on customer satisfaction, as did the handling of complaints [64]. Another result showed that higher customer satisfaction resulting in a positive influence on customer loyalty [64].

The researchers found that customers that had complaints in turn also had a greater customer satisfaction [64]. And customers who were satisfied would be loyal consumers [64]. Based on the analysis the researchers conducted.

5.6. Customer Complaint Handling on Customer Satisfaction

Another study was done on banks in Asia on the "effect of bank commitment, bank communication, and handling customer complaint on customer loyalty through customer satisfaction" was completed by obtaining a sample from customers [65]. "The commitment of a service provider greatly impacts on customer loyalty. In this case, when the service provider is highly committed to providing services to its customers, it will make customers satisfied and then will not switch to other competitors and continue to use the service or product" [65]. Similarly, as seen by the top three banks identified in this research project, the banks are heavily focused on providing exceptional service to their customers. The study found that commitment, communication and the way incidents are handled all have a great impact on one another [65].

In addition to the study pointing out that quality service will ensure customer satisfaction the study also indicated that bank commitment positively influence customer satisfaction [65]. Commitment is defined as "one of the most basic constructs of relationship used to satisfy customers" [65]. If there is a strong commitment, then it also drives long term relationships with customers [65]. Not only does bank commitment have a positive effect on customer satisfaction but also on loyalty which drives retention, which is what banks strive for with all of their customers [65].

A third aspect of the study done focused on the bank's communication and its impact on customer satisfaction. The research data showed that bank communication also had a positive and significant effect on customer satisfaction [65]. "Communication is very important in a business. The existence of good communication between service providers and their customers will provide satisfaction to these customers. [65] "Communication on this study is equivalent to receiving a timely response from the bank in the dataset used for this research project. Timely response was not one of the criteria taken into consideration for this research project since it lacked more information to truly relate it to the complaints received. Responses to the timely question in the dataset used were only listed as a 'yes' or 'no' answer. Yes and no are hard to quantify as an explanation for what constitutes

Figure 8. Word Cloud of Complaints Received

a timely response was not provided and could vary across the financial institutions in the dataset.

From the two million data records natural language processions identified the top complaints that companies received: top three being around loans, mortgages and payments(Figure 9). To identify areas of opportunities analysis of the top three financial institutions will be compared to the three financial institutions that have the least number of complaints against them. In addition to the bar chart in Figure 9 that shows the frequency of key words, the word cloud in Figure 8 clearly illustrates which words are most likely used in the complaints data.

The three companies that received the most complaints in our dataset against them were Equifax, Experian, and TransUnion regarding credit reporting. Since these three companies are not financial institutions and only serve as credit reporting bureaus they will be excluded as part of this analysis. The top three banks that received the most complaints after the credit reporting bureaus were Bank of America, Wells Fargo, and JP Morgan Chase. Understanding each financial institution first will allow for a comparison of the three companies and allow for analysis into what they have done in terms of responding to customer complaints and what they could do to decrease the number of complaints they receive.

5.7. Bank of America

Bank of America first came into the market in 1923 but was established as the leading bank with the merger between BankAmerica and NationsBank in 1998 [66]. Bank of America offers a large variety of products to their consumers, to include personal banking such as checking and
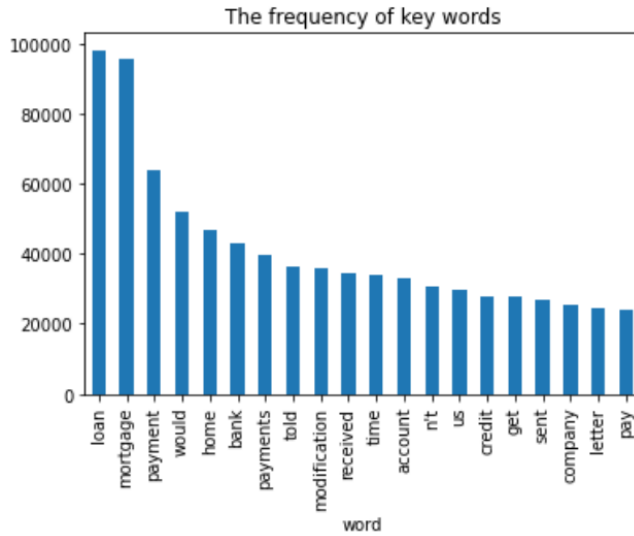
Figure 9. Frequency of Key Words

Figure 10. Comparison of Word Frequency between Bank of America, Wells Fargo, and JPMorgan Chase & Co.

savings accounts, home and personal loans, credit cards, small business accounts, wealth management activities, and products and support to larger businesses and institutions [61]. In 2019 Bank of America received the J.D. Power U.S. Retail Banking Advice Study highest ranking for customer satisfaction [67]. "The study measures satisfaction across twenty-three of the largest banks across the U.S." [67]. Bank of America received a score of 839 for their customer satisfaction. Comparing to our analysis, Bank of America should not be in the banks that receive the highestnumber of complaints given the high customer satisfaction rating for J.D. Power.

5.8. Wells Fargo

Wells Fargo was first founded in 1852 but similar to Bank of America current day Wells Fargo was established in 1998 with the merger of Wells Fargo & Company and Norwest Corporation [62]. Wells Fargo customers have many products to choose from including banking and credit cards, loans, investing and retirement support, wealth management, services to small businesses, and commercial products as well [68]. Wells Fargo is committed to transforming their business and practices, with a focus on customer experience and customer-focused innovation [69]. One of Wells Fargo's goals is customer service and advice and like Bank of America, Wells Fargo ranked third in customer satisfaction in the J.D. Power 2019 U.S. Retail Banking Advice Study [69].

5.9. JPMorgan Chase & Co.

Found in 2000, JPMorgan Chase & Co is the largest bank in the United States [70]. It offers an array of products for personal, business, and commercial clients. This includes checking and savings, credit cards, home loans, auto loans,

and a large investing branch [63]. In addition to J.D. Power rankings, banks are calculated net promoter scores which are calculated from responses provided by customers after they interact with the company's products or services [71]. Net promoter scores are a customer loyalty metric that "measures customers' willingness to not only return for another purchase or service but also make a recommendation to their family, friends or colleagues" [71]. JP Morgan's net promoter score is 8, which is considered good given that the average for major banks is around a zero.

5.10. Comparison of Top Three Companies with Highest Complaints

Initial word processing on the dataset, we were able to compare Bank of America, Wells Fargo, and JPMorgan Chase & Co. on the volume of complaints they received. As seen in Figure 10 the most seen words in customer's complaints for the three companies include loan, mortgage, and payments. Figure 11 further breaks the words most used in the complaints across the companies and looks to show how frequently it was used.

6. How analysis will help customer satisfaction

Understanding where customers complain the most will helps guide financial institutions towards improvements that will hopefully alleviate the pain points. A Bain & Company brief discusses the importance of analytics in deepening customer relationships in banking [72]. They outline that analytics "enable banks to activate, not just retain, their high-value customers" and harnessing the full power of analytics works best with five principles: "segmenting customers by value, automating forecasts, predicting loyalty, understanding what causes churn, and taking a test-and-learn approach" [72]. Our focus on complaints will help to inform why a customer might leave thus understanding customers' struggles. Our analysis of complaint data will lead to predicting

Figure 11. Comparison of Word Frequency between Bank of America, Wells Fargo, and JPMorgan Chase & Co.

Figure 12. "A Predictive NPS model yielded a sixfold increase in customer conversion for one bank" [72]

behavior of customers on certain products and issues. Bain & Company discusses how a predictive model using a data set with known drivers of Net Promoter Score (NPS) [72]. "A predictive model starts with basic features present in ordinary segmentation, such as channel usage, the frequency and nature of sales and service interactions, product usage, and revenue. It then taps more advanced sources of data, such as natural language processing of contact center conversations, including volume and tone [72]."

As seen in Figure 12, Bain & Company did a study on the impact of predictive NPS model and "using a predictive model achieves 70% predictive accuracy and a 30% success rate on commercial campaigns, compared with a roughly 5% success rate for the average campaign" [72]. Using the support of these successful number we will look to predict possible outcomes for Bank of America, Wells Fargo, and JPMorgan Chase & Co.

6.1. Complaints Handling on Bank Brand

In a study on "the impact of the magnitude of service failure and complaint handling on satisfaction and brand credibility in the banking industry" there is a correlation

Figure 13. Snippet of Code for creating Distribution graph

between customer satisfaction and how complaints are handled [73]. The study also points out the variability in what is considered "effective" handling; the fact that humans are involved introduces the variable that cannot be controlled [73]. An individual, despite what the bank may consider timely, can have differing opinions on how a complaint should be handled and what timeframe it should be handled in [73]. "Davidow (2003) concluded that there are six aspects of responsiveness: timeliness, redress, apology, credibility, attentiveness, and facilitation. Among these factors, timeliness is a controllable element that customers consider and judge firms regarding it since failure occurs" [73].

7. The Research

The research was conducted using machine learning and there were several graphs and matrixes produced. In the following sub-sections graphs generated will be discussed and a code snippet used to generate the graphs is shown as well.

7.1. Distribution Graphs

The histogram (code in Figure 13) is an acquainted graphical presentation for addressing the frequency of a batch of data. The scope of the information is divided into intervals and the number of qualities falling into each interval is counted. The histogram then, at that point, comprises of a progression of square shapes whose widths are characterized by as far as possible inferred by the bin widths, and whose statures rely upon the number of qualities in each bin.

In the above graphs (Figure 14), we are identifying the distribution graph for columns from the dataset which has no NAN values. The histograms are displayed with the frequency of counts in y-axis while x-axis contains the values of the column picked.

7.2. Correlation Graph & Matrix

The correlation graph (code in Figure 15) is a (K × K) square and even framework whose ij passage is the connection between the sections I and j of X. Huge data in this graph show genuine collinearity between the factors involved (Figure 16). In any case, the nonexistence
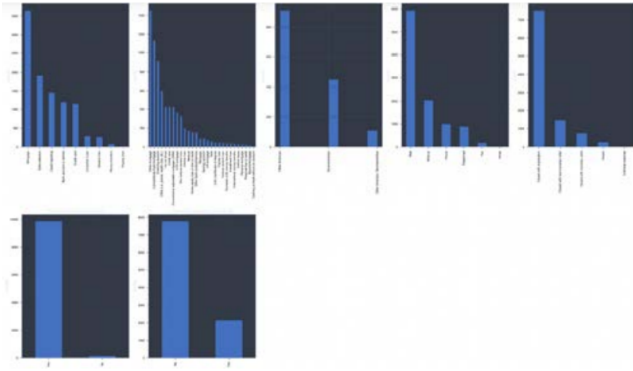
Figure 14. Distribution graphs

```
# Correlation matrix
def plotCorrelationMatrix(df, graphWidth):
    filename = df.dataframeName
    df = df.dropna('columns') # drop columns with NaN
    df = df[[col for col in df if df[col].nunique() > 1]] # keep columns where there are more than 1 unique values
    if df.shape[1] < 2:
        print(f'No correlation plots shown: The number of non-NaN or constant columns ({df.shape[1]}) is less than 2')
        return
    corr = df.corr()
    plt.figure(num=None, figsize=(graphWidth, graphWidth), dpi=80, facecolor='w', edgecolor='k')
    corrMat = plt.matshow(corr, fignum = 1)
    plt.xticks(range(len(corr.columns)), corr.columns, rotation=90)
    plt.yticks(range(len(corr.columns)), corr.columns)
    plt.gca().xaxis.tick_bottom()
    plt.colorbar(corrMat)
    plt.title(f'Correlation Matrix for {filename}', fontsize=15)
    plt.show()
```

Figure 15. Snippet of Code for creating Scatter and Density Plot

Figure 16. Correlation Matrix on Consumer Complaints dataset on Complaint ID

```
# Scatter and density plots
def plotScatterMatrix(df, plotSize, textSize):
    df = df.select_dtypes(include =[np.number]) # keep only numerical columns
    # Remove rows and columns that would lead to df being singular
    df = df.dropna('columns')
    df = df[[col for col in df if df[col].nunique() > 1]] # keep columns where there are more than 1 unique values
    columnNames = list(df)
    if len(columnNames) > 10: # reduce the number of columns for matrix inversion of kernel density plots
        columnNames = columnNames[:10]
    df = df[columnNames]
    ax = pady.plotting.scatter_matrix(df, alpha=0.75, figsize=[plotSize, plotSize], diagonal='kde')
    corrs = df.corr().values
    for i, j in zip(*plt.np.triu_indices_from(ax, k = 1)):
        ax[i, j].annotate('Corr. coef = %.3f' % corrs[i, j], (0.8, 0.2), xycoords='axes fraction', ha='center', va='cen
    plt.suptitle('Scatter and Density Plot')
    plt.show()
```

Figure 17. Snippet of Code for creating Scatter and Density Plot

of outrageous relationships doesn't suggest the absence of collinearity. The regressor factors for numerous relapses can be profoundly multicollinear even though no pairwise connections are huge. In the above dataset (Figure 16), we only have complaint-id as numerical with int64 data type. All other columns are objects. So, the graph is linear showing for only complaint-id.

7.3. Scatter Plot/Density Plot

A scatterplot (code in Figure 17) is quite possibly the most impressive yet basic visual plot accessible (Figure 18). In a scatterplot, the data points focused are set apart in Cartesian space with qualities of the dataset lined up with the directions. The credits are as a rule of nonstop information type. One of the critical perceptions that can be closed from a scatterplot is a connection between two ascribes under request. Assuming the qualities are directly connected, then, at that point, the information focuses adjust more like a fanciful straight line; in case they are not associated, the information focuses are dispersed. Aside from essential relationships, scatterplots can likewise demonstrate examples or gatherings of groups in the information and distinguish exceptions in the data.

8. Challenges & Opportunities

Getting access to data was straightforward and it was plentiful in quantity once downloaded. Access to customer complaints is made easy through the collection process of the Consumer Financial Protection Bureau. To really calculate customer satisfaction a deeper understanding of the

resolution would have been needed, such as the definition of what timely response means and whether the customer actually received a resolution to their complaint that did not require any further follow-up. This data would be gathered easily via a survey to customers. Companies use a CSAT, a customer satisfaction score, as a performance indicator for how their service is doing for certain products [74]. CSAT is measured through customer feedback and is done so by allowing customers a scale of one to five, ranging from Very Unsatisfied to Very Satisfied [74]. Once responses are collected, they are averaged out to provide a composite customer satisfaction score [74].

Another challenge to the research was having no demographic attributes to run analysis one in correlation with complaints. As seen in previous studies done on Indonesian banks, they were able to further extract insights based on age, gender, and education level. Such information could prove to be valuable to companies in identifying the root cause of the issue and adjusting their marketing towards those demographics that need it based on the data showed.

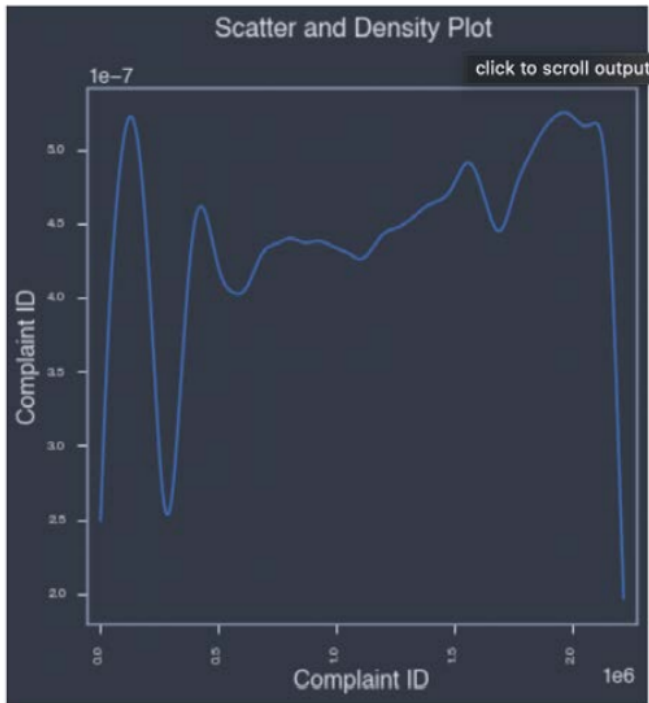To further develop this research topic, it would recommend collecting more data on the customers having

Figure 18. Scatter and Density Plot for Complaint ID

the problems. In addition to more data on the customers themselves, expanding the banks in this research project would be helpful in understanding if these issues are only limited to one company or whether there is an overarching problem with the industry.

9. Recommendations

The research has demonstrated that the top three banks to receive the most complaints were Bank of America, Wells Fargo, and JPMorgan Chase & Co. All three banks have extensive portfolios of products customers can chose from. Based on the data the most problematic areas for customers center around loans, mortgages, and payments. To address these top complaints banks need to try and find the true root of the problem by investigating whether there are any errors of the services on the side of the companies. Companies can draw out themes from gathering further explanations of the issues and identify where there are mistakes in service.

In addition to gathering additional details from customers on the true root cause of issues banks need to ensure they are continuing to improve their service response time, ensuring responses are still timely but quick and truly resolve the root problem for the customer. With timely responses also comes the need to actually to customer feedback and make improvements based on it. Banks should also strive to have frequent touchpoints with their customers across their various products. Customers that only have one product with the company should be reached out to and gather more understanding of why their portfolio with the company is not as robust as it could be. In addition,

customers that have multiple products from the company have great insight to how the company handles multiple items and can have feedback on consistency.

With the additional data, more analysis could be run to determine the true attributes that impact a positive customer satisfaction rate. Timely response would likely have a larger impact on satisfaction rates as customers want a quick turnaround when they have questions. Understanding the demographics can help inform the marketing teams to develop targeted messages. Understanding customers' portfolio size with the company could help pinpoint issues; do customers that only have a loan from a bank have more issues or is it customers that have multiple products experience more pain points, and are their issues with one or more of those products. Although this data would be highly useful gathering these details are a blocker for most research projects looking at gaining insight into multiple financial institutions. Most data in banks is considered confidential and considered as personally identifiable information (PII) making it difficult to obtain and run true cross comparison analysis on.

Acknowledgments

References

[1] K. States. (September 2, 2011) "three businesses that failed for lack of customer service". Accessed December 9, 2021. [Online]. Available: https://www.insidetucsonbusiness.com/news/on_guard/three-businesses-that-failed-for-lack-of-customer-service/article_42816a5e-d4df-11e0-9b64-001cc4c002e0.html

[2] R. F. Gerson, *Measuring customer satisfaction*. Menlo Park, Calif: Crisp Publications, 1993.

[3] Data World. "there are 802 open data datasets available on data.world.". Accessed December 9, 2021. [Online]. Available: https://data.world/datasets/open-data

[4] Consumer Financial Protection Bureau. "consumer complaint database.". Accessed December 9, 2021. [Online]. Available: https://www.consumerfinance.gov/data-research/consumer-complaints/

[5] S. Zad, M. Heidari, H. James Jr, and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *2021 IEEE World AI IoT Congress (AIIoT)*. IEEE, 2021, pp. 0255–0261.

[6] L. Cui and D. Lee, "Coaid: COVID-19 healthcare misinformation dataset," *CoRR*, vol. abs/2006.00885, 2020. [Online]. Available: https://arxiv.org/abs/2006.00885

[7] D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, and J. L. Zittrain, "The science of fake news," vol. 359, no. 6380, pp. 1094–1096, Mar. 2018. [Online]. Available: https://doi.org/10.1126/science.aao2998

[8] Q. Su, M. Wan, X. Liu, and C.-R. Huang, "Motivations, methods and metrics of misinformation detection: An NLP perspective," vol. 1, no. 1-2, p. 1, 2020. [Online]. Available: https://doi.org/10.2991/nlpr.d.200522.001

[9] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, April 3-7, 2017*, 2017, pp. 963–972. [Online]. Available: https://doi.org/10.1145/3041021.3055135

[10] S. Zad, M. Heidari, J. H. Jones, and O. Uzuner, "A survey on concept-level sentiment analysis techniques of textual data," in *2021 IEEE World AI IoT Congress (AIIoT)*. IEEE, 2021, pp. 0285–0291.

[11] M. Heidari, H. James Jr, and O. Uzuner, "An empirical study of machine learning algorithms for social media bot detection," in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. IEEE, 2021, pp. 1–5.

[12] C. Yang, R. C. Harkreader, and G. Gu, "Empirical evaluation and new design for fighting evolving twitter spammers," *IEEE Trans. Information Forensics and Security*, vol. 8, no. 8, pp. 1280–1293, 2013. [Online]. Available: https://doi.org/10.1109/TIFS.2013.2267732

[13] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds. Association for Computational Linguistics, 2019, pp. 4171–4186. [Online]. Available: https://doi.org/10.18653/v1/n19-1423

[14] M. Heidari and J. H. Jones, "Using bert to extract topic-independent sentiment features for social media bot detection," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, 2020, pp. 0542–0547.

[15] D. S. Khoury, D. Cromer, A. Reynaldi, T. E. Schlub, A. K. Wheatley, J. A. Juno, K. Subbarao, S. J. Kent, J. A. Triccas, and M. P. Davenport, "Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection," *Nature Medicine*, vol. 27, no. 7, pp. 1205–1211, May 2021. [Online]. Available: https://doi.org/10.1038/s41591-021-01377-8

[16] J. Havey. "pharma research progress hope,". [Online]. Available: https://catalyst.phrma.org/a-year-and-a-half-later-the-biopharmaceutical-industry-remains-committed-to-beating-covid-19?utm_campaign=2021-q3-cov-inn&utm_medium=pai_srh_cpc-ggl-adf&utm_source=ggl&utm_content=clk-pol-tpv_scl-geo_std-usa-dca-pai_srh_cpc-ggl-

[17] M. Heidari, S. Zad, B. Berlin, and S. Rafatirad, "Ontology creation model based on attention mechanism for a specific business domain," in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. IEEE, 2021, pp. 1–5.

[18] N. Sallahi, H. Park, F. E. Mellouhi, M. Rachdi, I. Ouassou, S. Belhaouari, A. Arredouani, and H. Bensmail, "Using unstated cases to correct for COVID-19 pandemic outbreak and its impact on easing the intervention for qatar," *Biology*, vol. 10, no. 6, p. 463, May 2021. [Online]. Available: https://doi.org/10.3390/biology10060463

[19] M. El-Harbawi, B. B. Samir, M.-R. Babaa, and M. I. A. Mutalib, "A new QSPR model for predicting the densities of ionic liquids," *Arabian Journal for Science and Engineering*, vol. 39, no. 9, pp. 6767–6775, Jun. 2014. [Online]. Available: https://doi.org/10.1007/s13369-014-1223-3

[20] P. R. Krause, T. R. Fleming, R. Peto, I. M. Longini, J. P. Figueroa, J. A. C. Sterne, A. Cravioto, H. Rees, J. P. T. Higgins, I. Boutron, H. Pan, M. F. Gruber, N. Arora, F. Kazi, R. Gaspar, S. Swaminathan, M. J. Ryan, and A.-M. Henao-Restrepo, "Considerations in boosting COVID-19 vaccine immune responses," *The Lancet*, vol. 398, no. 10308, pp. 1377–1380, Oct. 2021. [Online]. Available: https://doi.org/10.1016/s0140-6736(21)02046-8

[21] J. H. Kim, F. Marks, and J. D. Clemens, "Looking beyond COVID-19 vaccine phase 3 trials," *Nature Medicine*, vol. 27, no. 2, pp. 205–211, Jan. 2021. [Online]. Available: https://doi.org/10.1038/s41591-021-01230-y

[22] E. C. Fernández and L. Y. Zhu, "Racing to immunity: Journey to a COVID-19 vaccine and lessons for the future," *British Journal of Clinical Pharmacology*, vol. 87, no. 9, pp. 3408–3424, Jan. 2021. [Online]. Available: https://doi.org/10.1111/bcp.14686

[23] M. Heidari, S. Zad, and S. Rafatirad, "Ensemble of supervised and unsupervised learning models to predict a profitable business decision," in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. IEEE, 2021, pp. 1–6.

[24] Code Academy. "natural language processing/text preprocessing". [Online]. Available: https://www.codecademy.com/learn/dscp-natural-language-processing/modules/dscp-text-preprocessing/cheatsheet

[25] DeepAI. "named-entity recognition". [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/named-entity-recognition

[26] C. on foreign relations. "the u.s. war in afghanistan". [Accessed December 9, 2021]. [Online]. Available: https://www.cfr.org/timeline/us-war-afghanistan

[27] P. Hajibabaee, M. Malekzadeh, , M. Heidari, S. Zad, O. Uzuner, and J. H. Jones, "An empirical study of the graphsage and word2vec algorithms for graph multiclass classification," in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2021.

[28] "covid-19 third doses — health". [Accessed December 5, 2021]. [Online]. Available: https://www.fairfaxcounty.gov/health/novel-coronavirus/vaccine/third-doses

[29] S. Zad, M. Heidari, P. Hajibabaee, and M. Malekzadeh, "A survey of deep learning methods on semantic similarity and sentence modeling," in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2021.

[30] S. Akon and A. Bhuiyan, "Covid-19: Rumors and youth vulnerabilities in bangladesh," 07 2020.

[31] M. Heidari, S. Zad, M. Malekzadeh, P. Hajibabaee, , S. HekmatiAthar, O. Uzuner, and J. H. J. Jones, "Bert model for fake news detection based on social bot activities in the covid-19 pandemic," in *2021 12th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*. IEEE, 2021.

[32] "suicide statistics". [Accessed December 9, 2021]. [Online]. Available: https://www.befrienders.org/suicide-statistics

[33] T. V. Green and C. Doherty. "majority of u.s. public favors afghanistan troop withdrawal; biden criticized for his handling of situation". [Online]. Available: https://www.pewresearch.org/fact-tank/2021/08/31/majority-of-u-s-public-favors-afghanistan-troop-withdrawal-biden-criticized-for-his-handling-of-situation/

[34] M. U. Islam, F. B. Ashraf, A. I. Abir, and M. A. Mottalib, "Polarity detection of online news articles based on sentence structure and dynamic dictionary," in *2017 20th International Conference of Computer and Information Technology (ICCIT)*. IEEE, Dec. 2017. [Online]. Available: https://doi.org/10.1109/iccitechn.2017.8281777

[35] M. Malekzadeh, P. Hajibabaee, M. Heidari, S. Zad, O. Uzuner, and J. H. Jones, "Review of graph neural network in text classification," in *2021 12th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*. IEEE, 2021.

[36] E. Kiely and R. Farley. "timeline of u.s. withdrawal from afghanistan". [Online]. Available: https://www.factcheck.org/2021/08/timeline-of-u-s-withdrawal-from-afghanistan/

[37] Real Python. "tokenization in spacy". [Online]. Available: https://realpython.com/natural-language-processing-spacy-python\#tokenization-in-spacy

[38] D. Smeltz and E. Sullivan. (August 9, 2021) "us public supports withdrawal from afghanistan". [Online]. Available: https://www.thechicagocouncil.org/commentary-and-analysis/blogs/us-public-supports-withdrawal-afghanistan

[39] spaCy. "spacy 101: Everything you need to know". [Online]. Available: https://spacy.io/usage/spacy-101

[40] Code Academy. "text preprocessing". [Online]. Available: https://www.codecademy.com/courses/text-preprocessing/lessons/text-preprocessing/exercises/introduction

[41] J. Howard. "could covid-19 vaccine boosters be necessary? here's what experts are saying". [Accessed October 2021]. [Online]. Available: https://www.wesh.com/article/could-covid-19-vaccine-boosters-be-necessary-heres-what-experts-are-saying/36519793

[42] A. Ain, "The WHO is right to call a temporary halt to COVID vaccine boosters," *Nature*, vol. 596, no. 7872, pp. 317–317, Aug. 2021. [Online]. Available: https://doi.org/10.1038/d41586-021-02219-w

[43] E. Callaway, "COVID vaccine boosters: the most important questions," *Nature*, vol. 596, no. 7871, pp. 178–180, Aug. 2021. [Online]. Available: https://doi.org/10.1038/d41586-021-02158-6

[44] A. Weatherton. "health expert says booster shot could be needed after getting covid-19 vaccine". [Accessed June 8, 2021]. [Online]. Available: https://www.13newsnow.com/article/life/booster-shot-may-be-needed-after-covid-19-vaccine/291-49a8966c-3d91-48ad-99a0-02905c5593cc

[45] P. Naaber, L. Tserel, K. Kangro, E. Sepp, V. Jürjenson, A. Adamson, L. Haljasmägi, A. P. Rumm, R. Maruste, J. Kärner, J. M. Gerhold, A. Planken, M. Ustav, K. Kisand, and P. Peterson, "Dynamics of antibody response to BNT162b2 vaccine after six months: a longitudinal prospective study," *The Lancet Regional Health - Europe*, vol. 10, p. 100208, Nov. 2021. [Online]. Available: https://doi.org/10.1016/j.lanepe.2021.100208

[46] M. Heidari, J. H. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in *2020 International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2020, pp. 480–487.

[47] S. J. Thomas, E. D. Moreira, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart, J. L. Perez, G. P. Marc, F. P. Polack, C. Zerbini, R. Bailey, K. A. Swanson, X. Xu, S. Roychoudhury, K. Koury, S. Bouguermouh, W. V. Kalina, D. Cooper, R. W. Frenck, L. L. Hammitt, Özlem Türeci, H. Nell, A. Schaefer, S. Ünal, Q. Yang, P. Liberator, D. B. Tresnan, S. Mather, P. R. Dormitzer, U. Şahin, W. C. Gruber, and K. U. Jansen, "Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine through 6 months," *New England Journal of Medicine*, vol. 385, no. 19, pp. 1761–1773, Nov. 2021. [Online]. Available: https://doi.org/10.1056/nejmoa2110345

[48] E. Dolgin, "COVID vaccine immunity is waning — how much does that matter?" *Nature*, vol. 597, no. 7878, pp. 606–607, Sep. 2021. [Online]. Available: https://doi.org/10.1038/d41586-021-02532-4

[49] "virginia open data portal". [Accessed November 6, 2021]. [Online]. Available: https://data.virginia.gov/Government/VDH-COVID-19-PublicUseDataset-Vaccines-DosesAdmini/28k2-x2rj

[50] U.S. Food and Drug Administration. "covid-19 vaccines". [Accessed November 6, 2021]. [Online]. Available: https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines

[51] populationU. "populationU". [Accessed October 10, 2021]. [Online]. Available: https://www.populationu.com/us/virginia-population

[52] V. D. of Health. "covid-19 vaccine summary – coronavirus". [Accessed October 10, 2021]. [Online]. Available: https://www.vdh.virginia.gov/coronavirus/covid-19-in-virginia/covid-19-vaccine-summary/

[53] M. S. Berg. "what doctors wish patients knew about covid-19 herd immunity". [Accessed October 10, 2021]. [Online]. Available: https://www.ama-assn.org/delivering-care/public-health/what-doctors-wish-patients-knew-about-covid-19-herd-immunity

[54] D. Deshpande and M. Kumar, *Artificial Intelligence for Big Data - Complete guide to automating Big Data solutions using Artificial Intelligence techniques*. Packt Publishing, Limited, 2018.

[55] D. Lafrenière, *Delivering fantastic customer experienceFIX ME!!!!: how to turn customer satisfaction into customer relationships. 1st. edition.* New York: Productivity Press., 2019.

[56] A. Rao and S. Chandra, *The Little Book of Big Customer Satisfaction Measurement*. SAGE Publications India, 2013.

[57] Y. S. Maulana, D. Hadiani, D. S. Nurjanah, A. E. Mulyana, and A. N. Fajar, "An analysis of customer satisfaction and its effect on customer relationship management," in *Proceedings of the 1st Paris Van Java International Seminar on Health, Economics, Social Science and Humanities (PVJ-ISHESSH 2020)*. Atlantis Press, 2021. [Online]. Available: https://doi.org/10.2991/assehr.k.210304.004

[58] Consumer Financial Protection Bureau. "consumer response annual report". Accessed December 9, 2021. [Online]. Available: https://files.consumerfinance.gov/f/documents/cfpb_2020-consumer-response-annual-report_03-2021.pdf

[59] OCC. (April 6, 2019) "consumer protection". Accessed December 9, 2021. [Online]. Available: https://www.occ.gov/topics/consumers-and-communities/consumer-protection/index-consumer-protection.html

[60] G. Y. ERDURAN and F. LORCU, "THE INVESTIGATION OF ONLINE CUSTOMER COMPLAINTS IN THE BANKING SECTOR BY TEXT MINING," *Business & Management Studies: An International Journal*, vol. 8, no. 5, pp. 3835–3866, Dec. 2020. [Online]. Available: https://doi.org/10.15295/bmij.v8i5.1615

[61] "about bank of america - our people, our passion, our purpose.". Accessed December 9, 2021. [Online]. Available: https://about.bankofamerica.com/en

[62] Wikipedia. "wells fargo.". Accessed December 9, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Wells_Fargo

[63] "chase.com home" credit card, mortgage, banking, auto — chase online — chase.com. Accessed December 9, 2021. [Online]. Available: https://www.chase.com/

[64] A. Salim, M. Setiawan, R. Rofiaty, and F. Rohman, "Focusing on complaints handling for customer satisfaction and loyalty: The case of indonesian public banking," *EUROPEAN RESEARCH STUDIES JOURNAL*, vol. XXI, no. Issue 3, pp. 404–416, Nov. 2018. [Online]. Available: https://doi.org/10.35808/ersj/1071

[65] R. S. R. Widijanto and B. Rachmat, "Effect of bank commitment, bank communication and handling customer complaint on customer loyalty through customer satisfaction at PT bank central asia tbk of mojopahit mojokerto sub-branch office," *International Journal of Multicultural and Multireligious Understanding*, vol. 6, no. 3, p. 49, Jun. 2019. [Online]. Available: https://doi.org/10.18415/ijmmu.v6i2.756

[66] Wikipedia. "bank of america". Accessed December 9, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Bank_of_Americ

[67] newsroom. (February 4, 2019) "bank of america tops j.d. power ranking for retail banking advice.". Accessed December 9, 2021. [Online]. Available: https://newsroom.bankofamerica.com/press-releases/awards-and-recognition/bank-america-tops-jd-power-ranking-retail-banking-advice

[68] "wells fargo personal."wells fargo – banking, credit cards, loans, mortgages & more. Accessed December 9, 2021. [Online]. Available: https://www.wellsfargo.com/

[69] (2016) "wells fargo's transformation". Accessed December 9, 2021. [Online]. Available: https://www08.wellsfargomedia.com/assets/pdf/commitment/progress-report.pdf

[70] Wikipedia. "jpmorgan chase.". Accessed December 9, 2021. [Online]. Available: https://en.wikipedia.org/wiki/JPMorgan_Chase

[71] "j p morgan chase & co net promoter score 2021 benchmarks.". Accessed December 9, 2021. [Online]. Available: https://customer.guru/net-promoter-score/j-p-morgan-chase-co

[72] P. Baecker, M. Conde, D. Darnell, S. Narayanan, and M. Bergmann. (July 1, 2021) "how analytics can deepen banks' customer relationships". Accessed December 9, 2021. [Online]. Available: https://www.bain.com/insights/how-analytics-can-deepen-banks-customer-relationships/

[73] G. Shams, M. A. Rehman, S. Samad, and R. A. Rather, "The impact of the magnitude of service failure and complaint handling on satisfaction and brand credibility in the banking industry," *Journal of Financial Services Marketing*, vol. 25, no. 1-2, pp. 25–34, Mar. 2020. [Online]. Available: https://doi.org/10.1057/s41264-020-00070-0

[74] "what is csat and how do you measure it?". Accessed November 22, 2021. [Online]. Available: https://www.qualtrics.com/experience-management/customer/what-is-csat/