2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) | 978-1-6654-8684-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEMTRONICS55184.2022.9795855

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

Nan Jia
George Mason University

Lahari Bagam
George Mason University

Patricia Fabijanczyk
George Mason University

Ebrima Ceesay
George Mason University

***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**

I. 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]. Artificial intelligence can improve satisfaction rate in banking by using ML and NLP models [**?**], [2]–[7], [7]–[12] [13]–[20], [20]–[36].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. On the other hand, machine learning and deep learning approaches can help us to tackle problems in numerous scientific research fields ranging from financial industry and banking to structural engineering [37], [38] and medical image enhancement [39], [40].

A. 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

978-1-6654-8684-2/22/$31.00 ©2022 IEEE

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 [41]. Customer satisfaction counts for a lot today than all of the products a company can provide [41]. "If your customer is not satisfied, he or she will stop doing business with you [41]." Customer satisfaction is defined as the customer's "perception that his or her expectations have been met or surpassed [41]." 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 [41]. 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.

II. Dataset

The first dataset was sourced from DataWorld [42]. 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 [43]. 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 [43]. 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 [43]. 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.

A. Data Cleaning, Processing and Input Tools

As [44] proposed a modular experimental setting with the flow of data to results, 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.

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

TABLE I
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.

Text analysis with machine learning approaches has received significant attention over last decade [45]–[51]. Text analysis for investment and sentiment analysis of online comments can be use for improving user satisfaction in online banking [52]–[55] [56]. In this work, we implement business 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 [57]–[65]. Additionally, to compare the performance of our text classification models and define a baseline, we took the approach defined in [56]. Based on the accuracy and F1 score criteria, the new Transformer based language models are outperforming the traditional neural network models [56]. We utilize the BERT model as the main embedding layer for our NLP model. 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 business as well [66]–[74] in other domains such as health [39], [40], security and business, and apply transfer learning models [47] to analyse the online comments for creating a model for emergency response. From this data building out a decision tree such as paper [75] 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. Adding an additional layer to complaint analysis we are able to see that the sub-product that receives the most complaints is credit reporting. This is reflective of the data that was analyzed by the Consumer Financial Protection Bureau which also captured credit reporting as the product that received the most complaints. In addition, we see in Figure 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 complaints are mostly concentrated in the Eastern states of the U.S. with a significant number also originating from California and Texas (Figure 1). 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 manner.

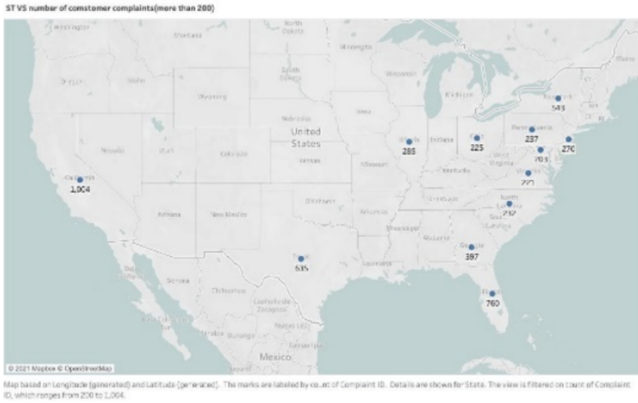

Fig. 1. States that received > 200 complaints

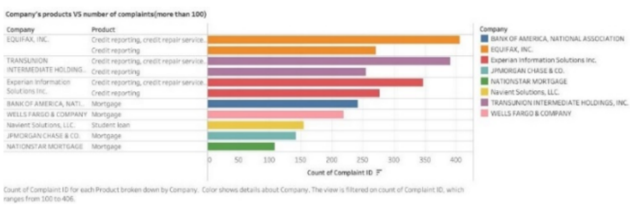

Fig. 2. Company's Products vs the Number of Complaints > 100

Different supervised machine learning techniques have been used in text classification [44]. Therefore, 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.

IV. 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 [76]. The machine power can process the dataset in order to calculate various types of variables from the population [76] 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 [43].

V. 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. [77]" 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 [77]. 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 [77]. Another staggering number is that 80% of customers would willingly pay more for a product or service if the customer experience is better [77]. 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 [77].

A. 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 [41]. 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 [41]. Surveys give companies the insights directly from their customers. "Strong relationships lead to higher levels of loyalty with customers [thus] resulting in profitability [78]". Service quality, when looking at the "Relationship Survey Framework", depends on the following customers experiences: installation, complaint, billing, purchase, pre-sales, and any other ones they interacted with [78]. 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 [78].

B. 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 [79]. "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" [79]. Similarly, as seen by the top three banks identified in this research project, the banks are heavily

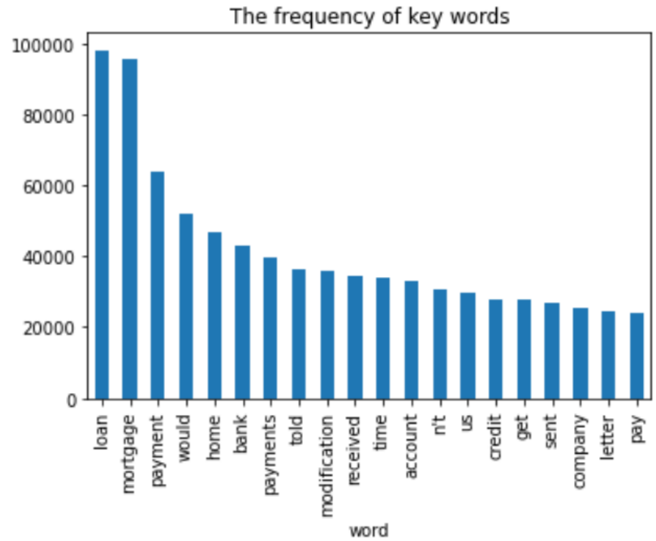

Fig. 3. Frequency of Key Words

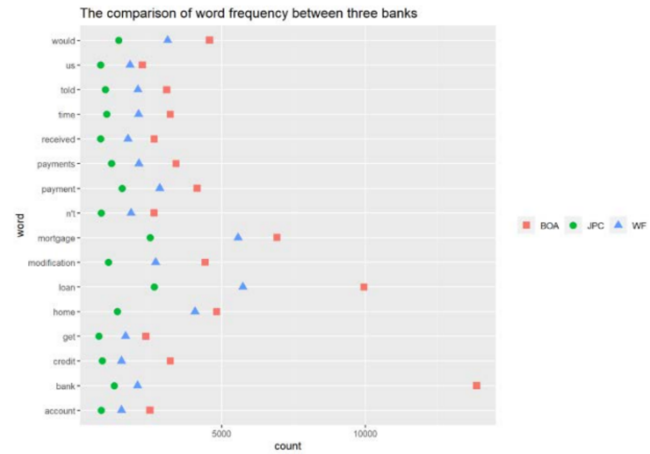

Fig. 4. Comparison of Word Frequency between Bank of America, Wells Fargo, and JPMorgan Chase & Co.

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 [79].

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 [79]. "Communication is very important in a business.

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 3). 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 3 that shows the frequency of key words.

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.

C. 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 [80]. Bank of America offers a large variety of products to their consumers, to include personal banking such as checking and savings accounts, home and personal loans, credit cards, small business accounts, wealth management activities, and products and support to larger businesses and institutions [81]. In 2019 Bank of America received the J.D. Power U.S. Retail Banking Advice Study highest ranking for customer satisfaction [82]. "The study measures satisfaction across twenty-three of the largest banks across the U.S." [82].

D. 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 [83]. 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 [84]. Wells Fargo is committed to transforming their business and practices, with a focus on customer experience and customer-focused innovation [85]. 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 [85].

E. 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 4 the most seen words in customer's complaints for the three companies include loan, mortgage, and payments. Figure 5 further breaks the words most used in the complaints across the companies and looks to show how frequently it was used.

VI. 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 [86]. They outline that analytics "enable banks to activate, not just retain, their high-value customers" and harnessing the full power of analytics

Fig. 5. Comparison of Word Frequency between Bank of America, Wells Fargo, and JPMorgan Chase & Co.

works best with five principles: "segmenting customers by value, automating forecasts, predicting loyalty, understanding what causes churn, and taking a test-and-learn approach" [86]. 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 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) [86]. "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 [86]."

The 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" [86]. Using the support of these successful number we will look to predict possible outcomes for Bank of America, Wells Fargo, and JPMorgan Chase & Co.

A. *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 between customer satisfaction and how complaints are handled [87]. 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 [87]. 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 [87]. "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" [87].

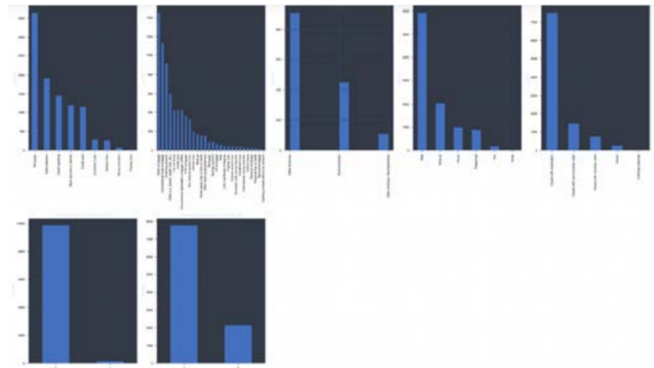

Fig. 6. Distribution graphs

VII. 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.

The histogram (code in Figure 6) 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 6), 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.

A. *Correlation Graph & Matrix*

The correlation graph (code in Figure 7) 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. In any case, the nonexistence 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, 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.

B. *Scatter Plot/Density Plot*

A scatterplot (code in Figure 7) is quite possibly the most impressive yet basic visual plot accessible (Figure 8). In a scatterplot, the data points focused are set apart in Cartesian

```
# 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='cer
    plt.suptitle('Scatter and Density Plot')
    plt.show()
```

Fig. 7. Snippet of Code for creating Scatter and Density Plot

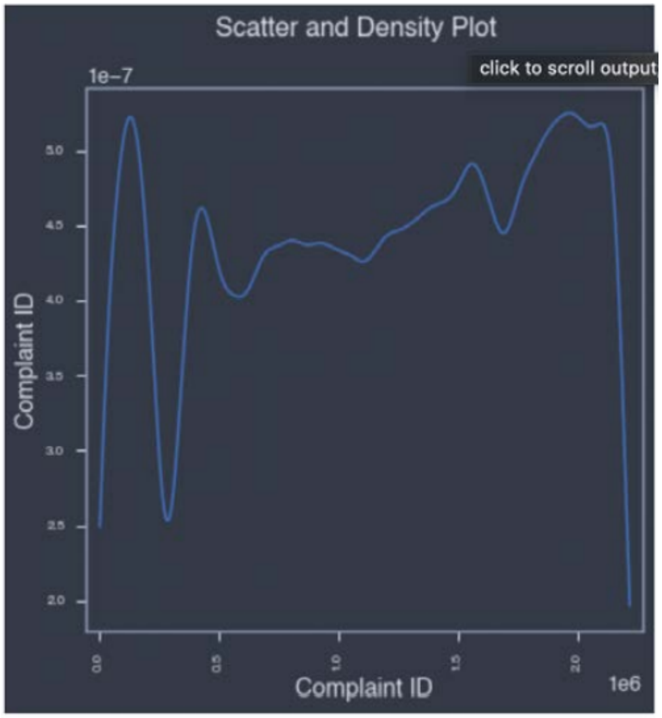

Fig. 8. Scatter and Density Plot for Complaint ID

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.

VIII. 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 [88]. 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 [88]. Once responses are collected, they are averaged out to provide a composite customer satisfaction score [88].

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 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.

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$_guard/three - businesses - that - failed - for - lack - of - customer - service/article_4 2816a5e - d4df - 11e0 - 9b64 - 001cc4c002e0.html$

[2] 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.

[3] G. D. S. Martino, S. Cresci, A. Barrón-Cedeño, S. Yu, R. D. Pietro, and P. Nakov, "A survey on computational propaganda detection," in *IJCAI*, 2020.

[4] J. Echeverrï£¡a, E. D. Cristofaro, N. Kourtellis, I. Leontiadis, G. Stringhini, and S. Zhou, "LOBO," in *Proceedings of the 34th Annual Computer Security Applications Conference*. ACM, Dec. 2018. [Online]. Available: https://doi.org/10.1145/3274694.3274738

[5] 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)*, 2020, pp. 480–487.

[6] M. Heidari and S. Rafatirad, "Semantic convolutional neural network model for safe business investment by using bert," in *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2020, pp. 1–6.

[7] X. Ruan, Z. Wu, H. Wang, and S. Jajodia, "Profiling online social behaviors for compromised account detection," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 176–187, 2016.

[8] M. Heidari, J. H. J. Jones, 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)*, 2021, pp. 1–5.

[9] J. Echeverría, E. D. Cristofaro, N. Kourtellis, I. Leontiadis, G. Stringhini, and S. Zhou, "Lobo – evaluation of generalization deficiencies in twitter bot classifiers," 2018.

[10] M. Heidari and S. Rafatirad, "Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment," in *2020 IEEE International Symposium on Technology and Society (ISTAS)*, 2020, pp. 322–329.

[11] M. Jiang, P. Cui, A. Beutel, C. Faloutsos, and S. Yang, "Inferring lockstep behavior from connectivity pattern in large graphs," *Knowledge and Information Systems*, vol. 48, 08 2016.

[12] M. Mazza, S. Cresci, M. Avvenuti, W. Quattrociocchi, and M. Tesconi, "Rtbust: Exploiting temporal patterns for botnet detection on twitter," 06 2019, pp. 183–192.

[13] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network model to recommend airline tickets by using online reviews," in *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA*, 2020, pp. 1–6.

[14] 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)*, 2021, pp. 0285–0291.

[15] 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 (IEMTRON-ICS)*, 2021, pp. 1–5.

[16] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Emergent properties, models, and laws of behavioral similarities within groups of twitter users," *Computer Communications*, vol. 150, pp. 47–61, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S014036641930283X

[17] 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)*, 2021, pp. 1–6.

[18] N. Chavoshi, H. Hamooni, and A. Mueen, "Debot: Twitter bot detection via warped correlation," *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 817–822, 2016.

[19] Wikipedia, "Sybil attack," https://en.wikipedia.org/wiki/Sybil_attack, 2021.

[20] J. Parsons, "Facebook's war continues against fake profiles and bots," https://www.huffpost.com/entry/facebooks-war-continues-against-fake-profiles-and-bots_b_6914282, 2021.

[21] M. Heidari, J. H. J. Jones, and O. Uzuner, "Online user profiling to detect social bots on twitter," 2022. [Online]. Available: https://arxiv.org/abs/2203.05966

[22] N. Z. Gong, M. Frank, and P. Mittal, "Sybilbelief: A semi-supervised learning approach for structure-based sybil detection," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 6, p. 976–987, Jun 2014. [Online]. Available: http://dx.doi.org/10.1109/TIFS.2014.2316975

[23] G. L. Caër, "A pearson random walk with steps of uniform orientation and dirichlet distributed lengths," *Journal of Statistical Physics*, vol. 140, no. 4, pp. 728–751, Jul. 2010. [Online]. Available: https://doi.org/10.1007/s10955-010-0015-8

[24] B. Carminati, E. Ferrari, and M. Viviani, "Security and trust in online social networks," in *Security and Trust in Online Social Networks*, 2013.

[25] 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

[26] 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)*, 2021, pp. 0515–0522.

[27] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, "Sybillimit: A near-optimal social network defense against sybil attacks," in *2008 IEEE Symposium on Security and Privacy (sp 2008)*, 2008, pp. 3–17.

[28] R. Abdolazimi, M. Heidari, A. Esmaeilzadeh, and H. Naderi, "Mapreduce preprocess of big graphs for rapid connected components detection," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 2022, pp. 0112–0118.

[29] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, pp. 25–39, 1983.

[30] S. Rafatirad and M. Heidari, "An exhaustive analysis of lazy vs. eager learning methods for real-estate property investment," 2019. [Online]. Available: https://openreview.net/forum?id=r1ge8sCqFX

[31] A. Mohaisen, A. Yun, and Y. Kim, "Measuring the mixing time of social graphs," in *Proceedings of the 10th annual conference on Internet measurement - IMC '10*. ACM Press, 2010. [Online]. Available: https://doi.org/10.1145/1879141.1879191

[32] M. Heidari and J. H. J. Jones, "Bert model for social media bot detection," 2022. [Online]. Available: http://hdl.handle.net/1920/12756

[33] Y. Xie, F. Yu, Q. Ke, M. Abadi, E. Gillum, K. Vitaldevaria, J. Walter, J. Huang, and Z. Mao, "Innocent by association: Early recognition of legitimate users," 10 2012.

[34] M. Heidari, "Nlp approach for social media bot detection(fake identity detection) to increase security and trust in online platforms," 2022.

[35] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove, "An analysis of social network-based sybil defenses," in *Proceedings of the ACM SIGCOMM 2010 conference on SIGCOMM - SIGCOMM '10*. ACM Press, 2010. [Online]. Available: https://doi.org/10.1145/1851182.1851226

[36] D. Mulamba, I. Ray, and I. Ray, "SybilRadar: A graph-structure based framework for sybil detection in on-line social networks," in *ICT Systems Security and Privacy Protection*. Springer International Publishing, 2016, pp. 179–193.

[37] P. Hajibabaee, F. Pourkamali-Anaraki, and M. Hariri-Ardebili, "Kernel matrix approximation on class-imbalanced data with an application to scientific simulation," *IEEE Access*, pp. 83 579–83 591, 2021.

[38] ——, "An empirical evaluation of the t-sne algorithm for data visualization in structural engineering," *arXiv preprint arXiv:2109.08795*, 2021.

[39] J. Liu, M. Malekzadeh, N. Mirian, T. Song, C. Liu, and J. Dutta, "Artificial intelligence-based image enhancement in pet imaging: Noise reduction and resolution enhancement," *PET clinics*, vol. 16, no. 4, pp. 553–576, 2021.

[40] M. Malekzadeh, T. Song, and J. Dutta, "Pet image denoising using unsupervised domain translation," in *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. IEEE, 2021.

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

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

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

[44] P. Hajibabaee, M. Malekzadeh, M. Ahmadi, M. Heidari, A. Esmaeilzadeh, R. Abdolazimi, and H. James Jr, "Offensive language detection on social media based on text classification," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022, pp. 0092–0098.

[45] 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, pp. 0466–0472.

[46] P. Hajibabaee, M. Malekzadeh, M. Heidari, S. Zad, O. Uzuner, and J. 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, pp. 0515–0522.

[47] M. Heidari, S. Zad, P. Hajibabaee, M. Malekzadeh, S. HekmatiAthar, O. Uzuner, and J. Jones, "Bert model for fake news detection based on social bot activities in the covid-19 pandemic," in *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2021, pp. 0103–0109.

[48] M. Malekzadeh, P. Hajibabaee, M. Heidari, S. Zad, O. Uzuner, and J. Jones, "Review of graph neural network in text classification," in *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2021, pp. 0084–0091.

[49] M. Malekzadeh, P. Hajibabaee, M. Heidari, and B. Berlin, "Review of deep learning methods for automated sleep staging," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022, pp. 0080–0086.

[50] A. Esmaeilzadeh, M. Heidari, R. Abdolazimi, P. Hajibabaee, and M. Malekzadeh, "Efficient large scale nlp feature engineering with apache spark," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022, pp. 0274–0280.

[51] J. Cummings, G. Lee, P. Nahed, M. E. Z. N. Kambar, K. Zhong, J. Fonseca, and K. Taghva, "Alzheimer's disease drug development pipeline: 2022," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 8, no. 1, p. e12295, 2022.

[52] 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

[53] 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

[54] 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

[55] 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

[56] A. Esmaeilzadeh and K. Taghva, "Text classification using neural network language model (nnlm) and bert: An empirical comparison," in *Proceedings of SAI Intelligent Systems Conference*. Springer, 2021, pp. 175–189.

[57] 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

[58] 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

[59] 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

[60] 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$_campaign = 2021 - q3 - cov - innutm_medium = pai_srh_cpc - ggl - adfutm_source = gglutm_content = clk - pol - tpv_scl - geo_std - usa - dca - pai_srh_cpc - ggl -$

[61] 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

[62] 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

[63] 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

[64] 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

[65] 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

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

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

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

[69] 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/

[70] 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

[71] 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/

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

[73] 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

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

[75] M. E. Z. N. Kambar, P. Nahed, J. R. F. Cacho, G. Lee, J. Cummings, and K. Taghva, "Clinical text classification of alzheimer's drugs' mechanism of action," in *Proceedings of Sixth International Congress on Information and Communication Technology*. Springer, 2022, pp. 513–521.

[76] 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.

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

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

[79] 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

[80] Wikipedia. "bank of america". Accessed December 9, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Bank$_of_Americ$

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

[82] 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

[83] Wikipedia. "wells fargo.". Accessed December 9, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Wells$_Fargo$

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

[85] (2016) "wells fargo's transformation". Accessed December 9, 2021. [Online]. Available: https://www08.wellsfargomedia.com/assets/pdf/commitment/progress-report.pdf

[86] 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/

[87] 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

[88] "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/