12-2023

PREDICTIVE MODEL FOR CFPB CONSUMER COMPLAINTS

Vyshnavi Nalluri

PREDICTIVE MODEL FOR CONSUMER FINANCIAL PROTECTION BUREAU CONSUMER COMPLAINTS

A Project

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment of the

Requirements for the Degree

Master of Science

in

Computer Science

by

Vyshnavi Nalluri

December 2023

PREDICTIVE MODEL FOR CONSUMER FINANCIAL PROTECTION

BUREAU CONSUMER COMPLAINTS

A Project

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Vyshnavi Nalluri

December 2023

Approved by:

Dr. Bilal Khan, Advisor, Computer Science and Engineering

Dr. Jennifer Jin, Committee Member

Dr. Ronald Salloum, Committee Member

ABSTRACT

Within the dynamic and highly competitive financial industry, the timely and efficient resolution of customer complaints stands as a central challenge, particularly in the intricate domain of mortgage services. The traditional processes for handling these complaints have long been recognized as laborious and resource-intensive, a situation that financial institutions, including the esteemed Wells Fargo, are keen to improve.

Currently, the industry largely relies on basic data analytics for identifying trends in customer complaints. However, this approach has its limitations, especially when dealing with complaints within the mortgage services domain. In response to this challenge, this research advocates the adoption of advanced predictive models as a groundbreaking solution. These models, powered by Random Forests hold the promise of transforming the management of mortgage-related complaints fundamentally.

The Random Forests model, known for its capacity to analyze complex, non-linear relationships within data, is poised to revolutionize the prediction of customer complaint resolution outcomes. By analyzing a vast dataset from the Consumer Complaint Database, comprising 3,211,591 complaints spanning a decade, the model aspires to see whether the mortgage-related complaints will be swiftly resolved or require an extended resolution time.

The anticipated outcomes of this endeavor encompass a transformative impact on the mortgage-related complaint resolution landscape.

While this research is a pivotal step forward, broader complaint categories, and further refined predictive models could enhance the efficacy of complaint management and resolution processes.

ACKNOWLEDGEMENTS

This research endeavor has been a journey of exploration, innovation, and growth, made possible through the support, guidance, and contributions of numerous individuals and institutions.

First and foremost, I would like to express my heartfelt appreciation to my research advisor Dr. Bilal Khan, Dr. Jennifer Jin, Dr. Ronald Salloum whose expertise, patience, and encouragement have been invaluable throughout this project. Your guidance in navigating the complexities of data analysis and machine learning has been pivotal in the development of the proposed solutions.

I would like to express my gratitude to the Consumer Financial Protection Bureau for granting me access, to their Consumer Complaint Database. This dataset played a role in my research allowing me to gain an understanding of consumer complaints within the financial sector. I am deeply grateful, for the efforts and support received from all the individuals and institutions involved. Although these acknowledgments only capture a portion of my appreciation, they truly reflect the impact each one of you has made on this research project.

TABLE CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

BACKGROUND

In today’s world the banking sector faces a hurdle; effectively and promptly addressing customer complaints. This challenge becomes more significant when it comes to complaints regarding mortgages which're an integral part of banking services. The significance of this matter is highlighted by the industry’s unwavering dedication to ensuring customer satisfaction and fostering loyalty.

Customer complaints do not represent grievances. Also offer valuable insights into an institution’s operational well-being and its relationships, with clients. The way these complaints are managed and resolved reflects an institution's dedication to providing exemplary service. This is especially critical in the context of mortgages, a linchpin of the financial sector.

Traditionally, many banks and financial institutions have leaned heavily on manual processes to investigate and resolve mortgage-related complaints. This labor-intensive approach often results in prolonged resolution times, negatively impacting operational efficiency and potentially eroding customer loyalty. The limitations of these conventional methods underline the urgent need for innovative, data-driven solutions.

1.1 Theoretical Framework

The theoretical foundation of this research rests on several key concepts which are Customer Satisfaction and Machine Learning & Predictive Analytics.

Customer Satisfaction is drawing from the extensive body of literature on customer satisfaction in the banking industry, this research builds on the understanding that satisfied customers are more likely to remain loyal to a financial institution and engage in positive word-of-mouth, which is essential for business growth (Anderson & Fornell, 1994; Oliver, 2014).

Machine Learning and Predictive Analytics:

Machine Learning and Predictive Analytics is a study that leverages machine learning techniques, particularly Random Forests, for predicting customer complaint resolution outcomes. The theoretical framework includes concepts from the field of predictive analytics, where historical data is used to make future predictions (Provost & Fawcett, 2013).

1.2 Statement of the Problem

The problem we aim to tackle is rooted in the inefficacy of conventional methods employed to manage mortgage-related complaints. These methods typically include basic data analytics and manual categorization approaches. For instance, basic data analytics may identify general trends within consumer complaints but often fall short in predicting future complaints or distinguishing between issues that can be promptly resolved and those requiring protracted resolution times. This inefficiency becomes pronounced in practice when financial institutions like Wells Fargo confront a substantial influx of mortgage-related complaints. The manual nature of these approaches leads to delays in addressing customer concerns, high operational costs, and inefficient allocation of resources, possibly resulting in customer dissatisfaction.

In the era of data-driven decision-making, we recognize that the banking sector has an opportunity to harness advanced analytics and machine learning for an enhanced approach to customer complaints. These modern methods offer the potential to streamline operations, bolster customer relations, and equip financial institutions to maintain a competitive edge. My approach, which employs advanced predictive models, seeks to address the limitations of traditional methods. It aspires to provide a comprehensive solution for effective complaint management in the mortgage services sector by significantly enhancing prediction accuracy and responsiveness, thus ensuring that financial entities like Wells Fargo can excel in addressing customer complaints in an efficient and cost-effective manner.

1.3 Hypothesis

Null Hypothesis (H0) is current manual complaint resolution methods are equally effective as machine learning-based predictive models in terms of timeliness and customer satisfaction.

Alternative Hypothesis (H1) is machine learning-based predictive models significantly outperform current manual complaint resolution methods in terms of timeliness and customer satisfaction.

1.4 Scope and Limitation

This study focuses primarily on mortgage-related complaints within Wells Fargo and similar financial institutions. The research employs machine learning techniques, specifically Random Forests, to predict customer complaint resolution outcomes. The analysis is grounded in the Consumer Complaint Dataset provided by the Consumer Financial Protection Bureau, which offers detailed information on individual consumer complaints, including complaint types, product categories, temporal trends, and geographic information.

Limitations include the reliance on historical data, which may not fully capture emerging trends, and the scope confined to mortgage-related complaints, which omits other areas of concern in the banking sector.

1.5 Conceptual Framework

The conceptual framework of this research encompasses the following key elements are Customer Complaint Resolution and Consumer Financial Protection Bureau (CFPB) Data.

Customer Complaint Resolution, At the core of this framework is the process of addressing and resolving customer complaints, which involves both manual and machine learning-based methods.

The research utilizes the extensive Consumer Complaint Dataset provided by the CFPB, which serves as the primary source of data for analysis.

1.6 Significance of the Study

This study carries substantial significance for both the banking industry and the broader field of data-driven decision-making. By developing predictive models for customer complaint resolution, this research contributes to enhancing operational efficiency, reducing costs, and improving customer satisfaction in the financial sector. It shows how an application of ML is addressed in real-world. Furthermore, it showcases how machine learning can be practically applied to tackle real world problems.

1.7 Explanation of Terminology

Resolving Customer Complaints. The procedure of handling and resolving customer grievances typically involving interactions, between customers and representatives from institutions.

Machine Learning. A branch of intelligence that enables systems to learn and enhance their performance based on experience without the need for programming.

Consumer Financial Protection Bureau (CFPB). An agency, within the United States government dedicated to safeguarding consumer interests in matters.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

In this chapter I thoroughly explore the knowledge regarding resolving customer complaints in the banking industry using machine learning techniques. The purpose of this literature review is to gain an understanding of the background that our research is based on.

2.2 Historical Perspective

Over the years the way the banking industry handles customer complaints has undergone changes. In the past there was an approach, to dealing with these complaints primarily because customers had limited options to express their concerns. However, as banking services expanded and communication methods advanced it became evident that a structured process, for resolving complaints was necessary (He et al., 2017).

2.3 Types of Customer Complaints

Today's banking institutions encounter a range of customer grievances that showcase the range of services they offer. Recent research has pinpointed recurring types of complaints such, as those pertaining to managing accounts, fees and charges fraudulent activities and issues related to mortgages (Fay et al., 2020). Among these concerns mortgage related complaints have garnered attention due to their nature and influence, on consumers (Consumer Financial Protection Bureau, 2020).

2.4 Impact on Financial Institutions

In years the banking industry has seen an increase, in the impact of customer complaints. Banks now recognize that unresolved issues can damage their reputation attract attention and even result in financial penalties (Altuntas et al., 2019). With growing competition in the banking sector institutions understand the importance of customer satisfaction and loyalty, for achieving long term success (Gaurav et al., 2021).

2.5 Regulatory Changes

Recent changes, in regulations have emphasized the significance of complaint handling, within the banking sector. For instance, in the United States the Consumer Financial Protection Bureau (CFPB) implemented rules mandating institutions to handle customer grievances systematically and promptly. As a result, banks have been prompted to invest in improved complaint resolution procedures and innovative technologies.

2.6 The Rise of Digital Banking

The rise of banking has significantly transformed the types of issues customers raise. As online and mobile banking options have expanded customers now possess an array of avenues to engage with their banks. However, this has also created opportunities for complaints to arise. Online reviews, social media platforms and specialized complaint channels have become channels for customers to express their concerns (Suki, 2016).

2.7 The Role of Data in Complaint Management

In today’s research it has been emphasized that data analytics plays a role, in handling customer complaints. Banks are now utilizing data to better understand the trends and patterns of complaints (Chen et al., 2021). This knowledge helps banks make decisions allowing them to take steps in resolving issues and enhancing customer service.

2.8 Challenges and Opportunities

Although there has been some improvement, in handling customer grievances within the banking sector obstacles remain. Recent research highlights difficulties like the requirement for resolution of complaints tackling the intricacies related to mortgage related issues and seamlessly integrating platforms into the resolution process (Gaurav et al., 2021; Verhoef et al., 2020). These challenges provide avenues, for innovation and data centric solutions.

2.9 Traditional Manual Approaches

In the past banks used to depend on methods, for examining and resolving customer complaints. This involved customer service agents investigating each issue separately which often led to procedures. However, these manual approaches were not efficient or scalable.

2.10 Technology-Driven Approaches

The use of data driven studies has made progress, in enhancing complaint resolution within the banking sector. However, these studies also have limitations. CRM systems, as discussed by Hendricks (2019) are effective in organizing and tracking complaints. Often lack the ability to predict how complaints will be resolved. This limitation can result in inefficient resource allocation. Longer resolution times. In my research I tackled this issue by introducing models that determine whether a mortgage related complaint will be resolved promptly thereby streamlining the overall complaint resolution process.

Moreover, while big data analytics have successfully predicted changes in complaint volumes as demonstrated by Davis et al. (2021) they may not fully capture the intricacies to mortgage services. In this project my goal was to address these limitations by customizing the model to account for the complexities in mortgage services. By doing I aimed to provide accurate predictions and improve the overall effectiveness of the complaint resolution process. Ultimately this approach seeks to redefine industry standards, from Wells Fargo's business perspective.

2.11 Data-Driven Resolution

Recent trends in data analytics have given rise to data-driven complaint resolution methods. Banks are increasingly using machine learning algorithms to analyze customer complaint data, identify patterns, and predict optimal resolution paths (Sharma et al., 2020). For instance, Natural Language Processing is employed to analyze customer complaint narratives, extracting sentiment and key issues (Zhu et al., 2018).

2.12 Real-Time Resolution

One of the emerging trends in complaint resolution is the move toward real-time solutions. Customers now expect swift responses to their complaints, particularly through digital channels. Chatbots and virtual assistants equipped with AI capabilities are being deployed to handle routine complaints promptly (Xu et al., 2020). These AI-driven systems try to provide immediate responses and guide customers through issue resolution steps.

Take, for example the fashion retailer H&M. They employ chatbots, in their customer service operations to help customers with asked questions regarding product availability order status and returns. By having these chatbots handle queries H&Ms human customer service agents can focus on intricate and tailored interactions, with customers ultimately enriching the overall shopping experience.

2.13 Regulatory Compliance

Regulatory authorities have had an impact, on shaping the ways in which complaints are resolved. Banks are obligated to follow guidelines set forth by regulators, which emphasize the importance of resolving complaints transparently (Consumer Financial Protection Bureau, 2013). Noncompliance, with these regulations can lead to fines and harm to one’s reputation.

2.14 Customer Self-Service

To meet the growing need, for self service solutions banks are working on creating customer portals and mobile applications that enable customers to address concerns, on their own. This strategy does not improve customer satisfaction. Also alleviates the workload of customer service teams (Hendricks, 2019).

2.15 Challenges and Innovations

Despite the progress made in resolving complaints there are still challenges. These challenges involve the integration of platforms ensuring the privacy and security of data and dealing with the intricacies of complaints related to mortgages. Researchers are currently investigating solutions such, as Artificial Intelligence and predictive analytics to tackle these issues (Sharma et al., 2020).

2.16 The Role of Machine Learning

The use of machine learning, in customer service has grown considerably in the years. It has become a tool, for analyzing and deriving information from vast amounts of data, which helps organizations make informed decisions based on data. In terms of handling customer complaints machine learning streamlines. Provides capabilities revolutionizing how companies deal with customer issues (SAS, 2021).

2.17 Applications in Customer Service

Machine learning models of today can analyze customer feedback reviews and social media data to determine sentiment. This enables banks to gauge customer sentiment, identify areas of concern, and respond promptly to negative feedback (Liu, 2020).

AI-powered chatbots equipped with machine learning capabilities are now commonplace in customer service. These chatbots provide instant responses, answer inquiries, and even handle routine complaint resolution tasks. A recent survey by Gartner predicts that by 2025, 70% of customer interactions will involve emerging technologies such as chatbots (Gartner, 2021).

2.18 Success Stories and Benefits

AI driven chatbots and automated systems have the advantage of providing round the clock responses, which significantly speeds up response times. According to a study conducted by Microsoft their chatbot managed to reduce response times by, then 25% (Microsoft, 2021).

When it comes to customer satisfaction, personalized recommendations and tailored responses powered by machine learning play a role. A study conducted by McKinsey & Company highlights how customized customer experiences can lead to satisfaction scores increasing by 10 15% (McKinsey, 2021).

By implementing automation through machine learning organizations do not lighten the workload for agents but also achieve cost savings. Additionally, it helps eliminate errors and redundancies, in customer service processes as demonstrated in a study carried out by Deloitte (Deloitte, 2021).

2.19 The Consumer Complaint Dataset

The study relies on the Consumer Complaint Database, which is made available by the Consumer Financial Protection Bureau (CFPB). This dataset is a resource, for analyzing consumer complaints in the sector. It provides information on aspects of these complaints making it an excellent source for analysis purposes.

The CFPB, established in 2010 as part of the Dodd Frank Wall Street Reform and Consumer Protection Act has played a role in safeguarding consumers interests, within the industry. They. Regularly update this dataset to ensure transparency and accountability in this sector.

2.20 Data Characteristics and Granularity

The dataset provided by the CFPB spans from 2012 to 2022. Consists of 3,211,591 records. These records contain all the information we need.

Each entry, in the dataset includes the following details:

Date Received: The date when the complaint was received by the CFPB.

Product – Describes the type of product involved.

Sub-Product – Specifies the subcategory of the product.

Issue – Explains what specific problem occurred with that product.

Sub-issue – Provides details, about any sub issue related to the problem.

Consumer Complaint Narrative – A written description of the consumers experience.

Company – The name of the institution involved.

State – The state where the problem occurred.

Zip Code – The zip code corresponding to that state.

Tags – A unique number assigned to each complaint.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

The approach used to tackle the research issue of forecasting customer complaint resolution outcomes and complaint counts in the banking sector involved two parts: creating predictive models through machine learning.

3.2 Research Design Quantitative Research Approach

In this study a quantitative approach is utilized to examine the Consumer Complaint Dataset and construct models. Quantitative research is well suited for this type of analysis as it enables the examination of datasets and facilitates the generation of significant predictions (Creswell & Creswell 2017).

3.3 Consumer Complaint Database

The main data source utilized for this study is the Consumer Complaint Database, which is made available by the Consumer Financial Protection Bureau (CFPB). This dataset can be accessed by the public. Offers information regarding consumer complaints. It includes data, on complaint types, product categories, temporal patterns, and geographic specifics.

Furthermore, it contains information, in the form of consumer complaint narratives that offer valuable insights into individual grievances.

3.4 Data Pre-processing

Data preprocessing plays a role, in the research methodology as it focuses on getting the Consumer Complaint Dataset ready, for analysis. This phase involves operations to clean, transform and structure the dataset. The subsequent sections will provide explanations of the data preprocessing procedures used in this study:

3.5 Data Loading

To start we load the dataset from a file path by utilizing the read_csv function from the Pandas library. To get an understanding of the datas structure we display a preview of the few rows of the dataset.

3.6 Date Conversion

To conduct time-based analysis, we convert the "Date received" and "Date sent to company" columns into a format. This conversion guarantees that we can perform operations related to dates.

3.7 Average Resolution Time Calculation

One important measure we focus on is the duration it takes to address each consumer complaint. To determine this, we calculate the time gap, between the "Date sent to company". Date received" columns. We then create a column called "Days, between " where we store these calculated values. A new column, "Days between," is created to store these calculated values. Subsequently, the .mean() method is applied to determine the average resolution time for all complaints in the dataset.

3.8 Column Removal

To streamline the dataset and focus on the most relevant information, several columns, including "Consumer complaint narrative," "Company public response," and "Tags," are removed from the dataset.

3.9 Filtering by Company

As the scope of this research is primarily centered on Wells Fargo, the dataset is refined by filtering for complaints specifically related to "WELLS FARGO & COMPANY." This step narrows the analysis down to this specific financial institution, aligning with the primary focus of the project.

3.10 Recalculation of Average Resolution Time (Wells Fargo)

Following the filtering process, the code recalculates the average number of days taken to resolve complaints, with a particular focus on those concerning Wells Fargo. This specific evaluation provides insights into the complaint resolution times within the chosen financial institution.

By implementing these data preprocessing steps, the dataset is prepared for subsequent analysis, including machine learning, predictive modeling. This rigorous preparation ensures that the dataset is clean, structured, and conducive to addressing the research problem effectively.

3.11 Binary Encoding of Target Variable

Next, involved the transformation of the target variable 'Consumer disputed?' into binary format. This transformation was accomplished by mapping 'No' to 0 and 'Yes' to 1. By converting the target variable into a binary classification problem, the research focuses on predicting whether a consumer dispute will be raised (1) or not (0).

3.12 Handling Categorical Columns

A critical aspect of data preprocessing was the management of categorical columns within the dataset. The following categorical columns were identified: 'Product,' 'Sub-product,' 'Issue,' 'Sub-issue,' 'Company,' 'State,' 'ZIP code,' 'Consumer consent provided?’, 'Submitted via,' 'Date sent to company,' 'Company response to consumer,' 'Timely response?', 'Consumer disputed?', 'Complaint ID,' and 'Days between.'

To address missing values within these categorical columns, the approach was to replace them with the 'Unknown' category. This ensured data completeness and allowed for effective analysis and modeling.

3.13 Label Encoding

To facilitate the application of machine learning algorithms, label encoding was applied to the categorical data. This process involved converting categorical values into numerical labels, making them suitable for machine learning model training. For each of the identified categorical columns, a Label Encoder was applied. It was imperative to confirm that each column was of string data type to prevent errors during the encoding process.

3.14 Handling Class Imbalance

The dataset exhibited class imbalance, particularly in the 'Consumer disputed?' variable. This variable had three classes: 'Yes,' 'No,' and 'Unknown.' To tackle this imbalance and guarantee the dependability of the models we utilized a technique called Synthetic Minority Over Sampling Technique (SMOTE). SMOTE helps address the class referred to as 'Yes,', by generating samples. This approach rebalances the dataset ensuring that the predictive models don't show bias towards the majority class. As a result, it provides predictions for consumer dispute outcomes. By implementing these steps in data processing, we are successful in preparing the dataset, for training and evaluating the models.

3.15 Predictive Models Approach

To predict the outcomes of customer complaint resolutions we utilize machine learning techniques. The goal is to create models that can accurately differentiate between complaints that are resolved quickly and those that take time to be resolved. These predictions are crucial, for institutions like Wells Fargo as they help anticipate changes, in the number of complaints received.

3.16 Machine Learning Models

In the field of machine learning Random Forests are commonly used for modeling. They are known for their flexibility, in handling datasets capturing patterns and delivering accurate predictions. They are well-suited for addressing the research problem, which involves predicting customer complaint resolution outcomes based on a wide array of features.

3.17 Model Development

Two distinct approaches were employed for the prediction of customer complaint resolution outcomes. These approaches are detailed below:

For the prediction of customer complaint resolution outcomes, I chose a Random Forest classifier as the machine learning model of choice. I divided the dataset into training (80%) and testing (20%) subsets using the train_test_split function from the scikit-learn library. I initialized the RandomForestClassifier from scikit-learn and trained the model on the training data. Subsequently, I made predictions on the testing data using the trained model. I evaluated the model's performance based on several metrics:

- Accuracy - This metric measures the overall correctness of my model's predictions concerning all classes.
- Precision - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The macro average considers precision for each class and provides a metric that is not dominated by the majority class.
- Recall - Recall measures the ratio of correctly predicted positive observations to all actual positive observations. As with precision, I used the macro average.
- F1 Score - The F1 score is the harmonic mean of precision and recall. Like precision and recall, I employed the macro average.

I evaluated the model's performance, as detailed by these metrics, to determine its effectiveness in predicting customer complaint resolution outcomes. The Random Forest model's results play a crucial role in assessing my model's predictive capabilities for timely complaint resolution.

CHAPTER FOUR

DATA ANALYSIS AND RESULTS

4.1 Exploratory Data Analysis

The exploratory data analysis (EDA) conducted in this research is a critical component of understanding the dataset's characteristics and deriving initial insights. EDA serves as a foundation for more advanced analysis and is instrumental in uncovering trends, patterns, and outliers in the data. In this section, I will analyze the key findings from the EDA and provide insights on expected visualizations that can enhance the understanding of the data.

4.2 Calculating the Average Number of Days Taken to Resolve Issues

The calculated average of approximately 1.80 days to resolve consumer complaints provides an initial insight into the efficiency of the company's issue resolution process.

4.3 Temporal Patterns

The dataset encompasses a substantial time frame, spanning from January 2012 to December 2022, with monthly records. Several noteworthy observations can be drawn. Firstly, it is evident that the number of consumer complaints is subject to temporal variations. Throughout this decade-long period, there are noticeable fluctuations, and specific time frames exhibit spikes in complaint volumes.

- 2012 to 2015 - During this period, there's a steady increase in complaints, with some seasonal fluctuations.

- 2015 to 2017 - Complaints continue to increase but with more significant fluctuations. Hence, indicate evolving customer sentiment or regulatory changes impacting complaint volumes.
- 2017 to 2019 - Complaints start to decrease.
- 2019 to 2020 - A slight decrease in complaints is noted.
- 2021 to 2022 - Some increase, with spikes in late 2021 and a significant increase in December 2022.

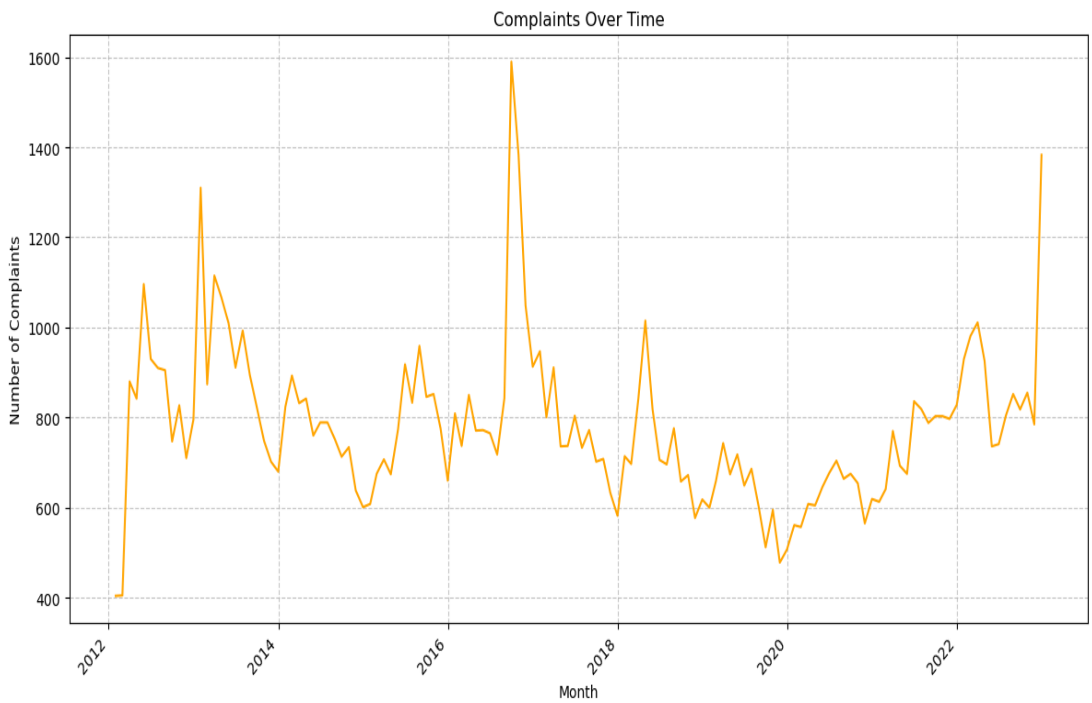

Figure 1 Complaints Over Time

4.4 Complaints by Product Type

The analysis of complaints by product type reveals significant insights into the areas of Wells Fargo's operations that generate the most consumer grievances. At the forefront, mortgage-related issues stand out, with a substantial 42,543 complaints. This signals that the mortgage department requires careful attention, as these complaints may indicate potential challenges in this service area.

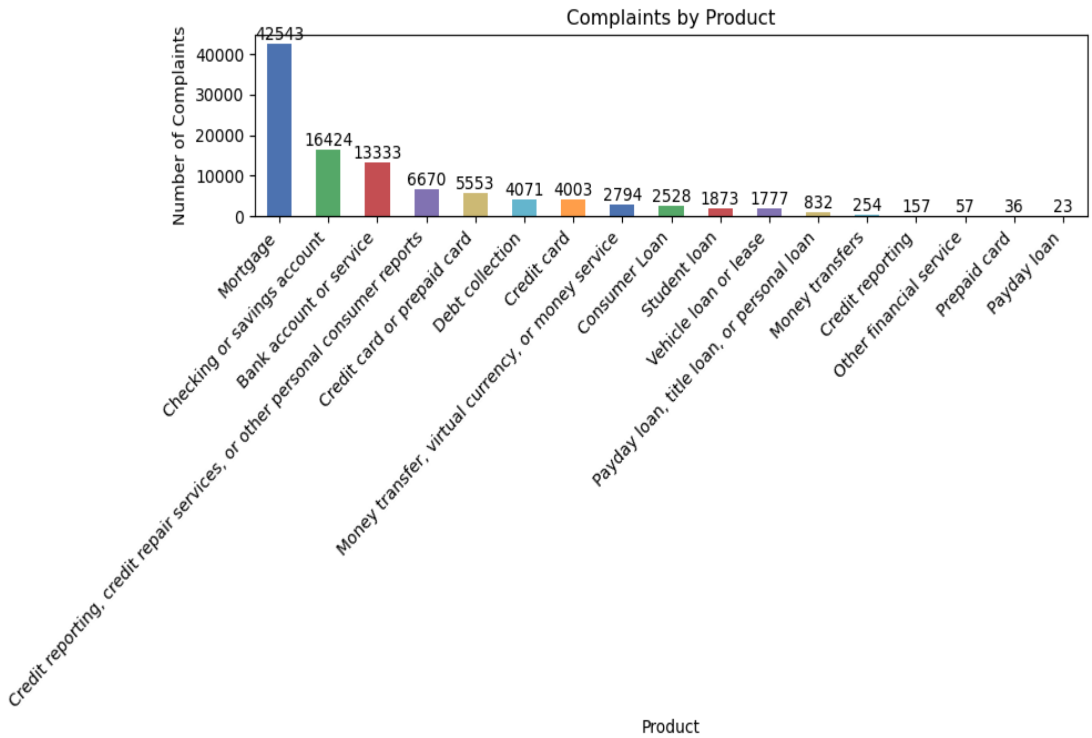

Figure 2. Complaints by Product

Following closely are complaints related to checking or savings accounts, with 16,424 reports. These encompass a range of issues, including fees, unauthorized transactions, and account management problems. Bank accounts and services come next, with 13,333 complaints, reflecting a broader set of concerns within the general banking sphere. Notably, the credit reporting, credit repair services, or other personal consumer reports category received 6,670 complaints, emphasizing potential issues with credit-related services. Credit cards, debt collection, and other services have also triggered numerous complaints. These findings are essential for Wells Fargo as they pinpoint areas that require focused improvement, resource allocation, and customer service adjustments.

4.5 Common Issues

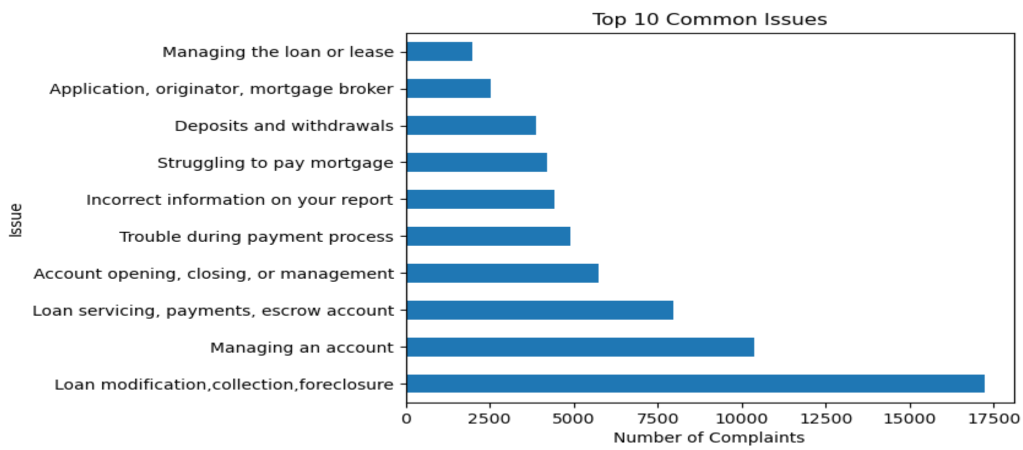

Figure 3. Top 10 Common Issues

The analysis of the top 10 common issues reported by Wells Fargo customers sheds light on the most prevalent challenges faced by consumers. Leading the list is the issue of "Loan modification, collection, foreclosure," with a substantial count of 17,241 complaints. This category encompasses concerns related to loan modifications, debt collection, and foreclosure processes, indicating potential complexities in these areas of Wells Fargo's services.

Following closely is "Managing an account" with 10,368 complaints, emphasizing challenges in account management, including issues like fees, unauthorized transactions, and general account-related problems. "Loan servicing, payments, escrow account" is another significant concern, with 7,967 complaints, suggesting potential difficulties in loan servicing and payment processes, including escrow management.

"Account opening, closing, or management" received 5,728 complaints, highlighting potential issues with the onboarding and offboarding processes for customers. "Trouble during the payment process" is the fifth most common issue, with 4,913 complaints, signifying payment-related challenges that customers encounter.

"Incorrect information on your report" follows with 4,437 complaints, indicating inaccuracies in credit reports. "Struggling to pay the mortgage" is another pressing issue, with 4,227 complaints, suggesting that some customers face financial difficulties in meeting their mortgage obligations.

"Deposits and withdrawals" received 3,870 complaints, indicating issues related to banking transactions. "Application, originator, mortgage broker" is the ninth most common problem, with 2,542 complaints, possibly relating to difficulties in the application process or with third-party mortgage brokers. Lastly, "Managing the loan or lease" received 1,979 complaints, highlighting concerns in overseeing existing loans or leases. This analysis shows us areas of improvement needed to be done.

4.6 Complaints by State

The way complaints are distributed across states offers insights, into the geographical patterns of customer problems related to Wells Fargo. It's worth noting that California (CA) has the number of complaints with a count of 18,898 making it a focal point for customer concerns. Following behind is Florida (FL), with 10,506 complaints, which indicates challenges faced by Wells Fargo customers in that state. Texas (TX) is another state where there’s several complaints totaling 7,106.

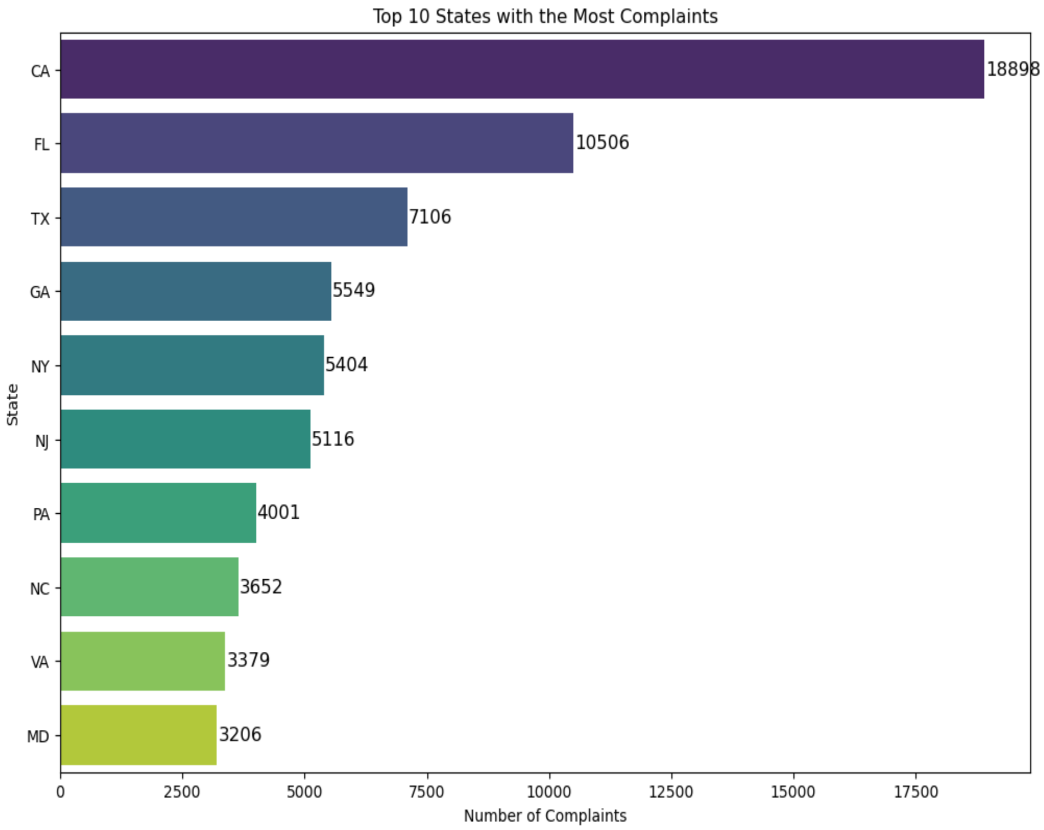

Figure 4. Top 10 States with Complaints

Georgia (GA) and New York (NY) are also quite notable having received 5,549 and 5,404 complaints respectively. This suggests that customers, in these states face several issues when it comes to the services provided by Wells Fargo. Similarly, New Jersey (NJ) and Pennsylvania (PA) report 5,116 and 4,001 complaints respectively highlighting the prevalence of concerns in these regions.

North Carolina (NC) and Virginia (VA) on the hand each have over 3,000 complaints, with figures standing at 3,652 and 3,379 respectively. Maryland (MD) and Arizona (AZ) each have over 3,000 complaints as well, with 3,206 and 2,777, respectively.

While these states demonstrate higher complaint volumes, it's essential to note that some complaints are marked as "None," totaling 2,193. These could be instances where the complainant's state wasn't specified.

4.7 Count of Common Sub-Issues for Mortgage Complaints

Table 1. Count of Common Sub-issues for Mortgage Complaints

Count of Common Sub-Issues for Mortgage Complaints	
Loan modification, collection, foreclosure	17241
Loan servicing, payments, escrow account	7967
Trouble during payment process	4913
Struggling to pay mortgage	4227
Application, originator, mortgage broker	2542
Applying for a mortgage or refinancing an existing mortgage	1753
Settlement process and costs	1078
Credit decision / Underwriting	854
Closing on a mortgage	802
Other	720

Mortgage-related complaints received by Wells Fargo reveal several prevalent sub-issues that customers have encountered. Topping the list is "Loan Modification, Collection, Foreclosure," with a significant 17,241 complaints. This sub-issue encompasses various challenges associated with loan modifications, debt collection, and foreclosure processes, indicating a substantial area of concern for customers. Following closely is the sub-issue of "Loan Servicing, Payments, Escrow Account," with 7,967 complaints.

This suggests that many customers have experienced difficulties in managing their loans, payments, and escrow accounts. "Trouble During Payment Process" is the third most common sub-issue, with 4,913 complaints, indicating struggles that customers face during the payment process, such as payment processing errors. "Struggling to Pay Mortgage" follows, with 4,227 complaints, reflecting the challenges some customers encounter in meeting their mortgage payment obligations. "Application, Originator, Mortgage Broker" is also a notable concern, with 2,542 complaints, suggesting issues with mortgage applications and originators. Additionally, customers have raised 1,753 complaints regarding the process of applying for a new mortgage or refinancing an existing one. Furthermore, the "Settlement Process and Costs" and "Credit Decision / Underwriting" sub-issues have received 1,078 and 854 complaints, respectively.

These numbers indicate customer dissatisfaction with the settlement process, particularly concerning costs, and challenges related to credit decisions and underwriting. Lastly, 802 complaints pertain to issues during the "Closing on a Mortgage" process, which is a critical final step in obtaining a mortgage. There is also a category labeled "Other" that comprises 720 complaints. While this category is less specific, it highlights that some complaints do not neatly fit into the predefined sub-issues.

4.8 Top Mortgage Complaints by State

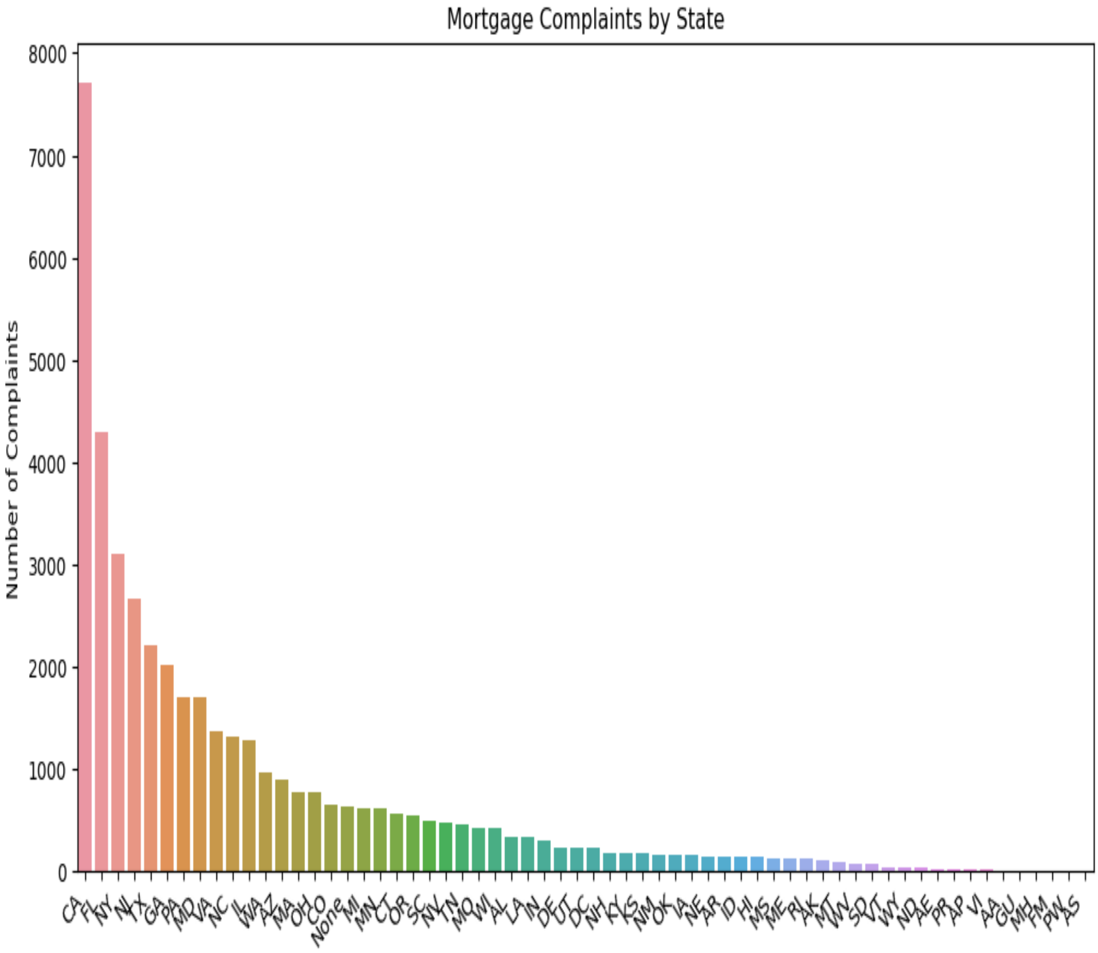

Figure 5. Top Mortgage Complaints by State

Table 2. Count of Mortgage Complaints by State

Count of Mortgage Complaints by State	
CA	7713
FL	4298
NY	3094
NJ	2660
TX	2213
GA	2007
PA	1695
MD	1690
VA	1358
NC	1309

The distribution of mortgage-related complaints at Wells Fargo by state reveals important geographic insights into customer experiences. California (CA) has the highest number of mortgage complaints, with 7,713 cases. This is a significant concern and suggests that mortgage-related issues are prominent in this state. Florida (FL) follows with 4,298 complaints, highlighting that customers in Florida also face challenges in their mortgage interactions with Wells Fargo. New York (NY) comes in third with 3,094 complaints, signifying that a considerable number of customers in this state have raised concerns about their mortgage experiences. New Jersey (NJ) and Texas (TX) are fourth and fifth on the list, with 2,660 and 2,213 complaints, respectively.

These numbers emphasize the regional variation in customer experiences and concerns.

Georgia (GA) and Pennsylvania (PA) both report around 2,000 complaints, with 2,007 and 1,695, respectively. Maryland (MD) and Virginia (VA) follow closely with 1,690 and 1,358 complaints, indicating that mortgage issues are significant in these states as well. North Carolina (NC) rounds out the list of the top ten states for mortgage-related complaints, with 1,309 cases.

4.9 Mortgage Complaints Over Time

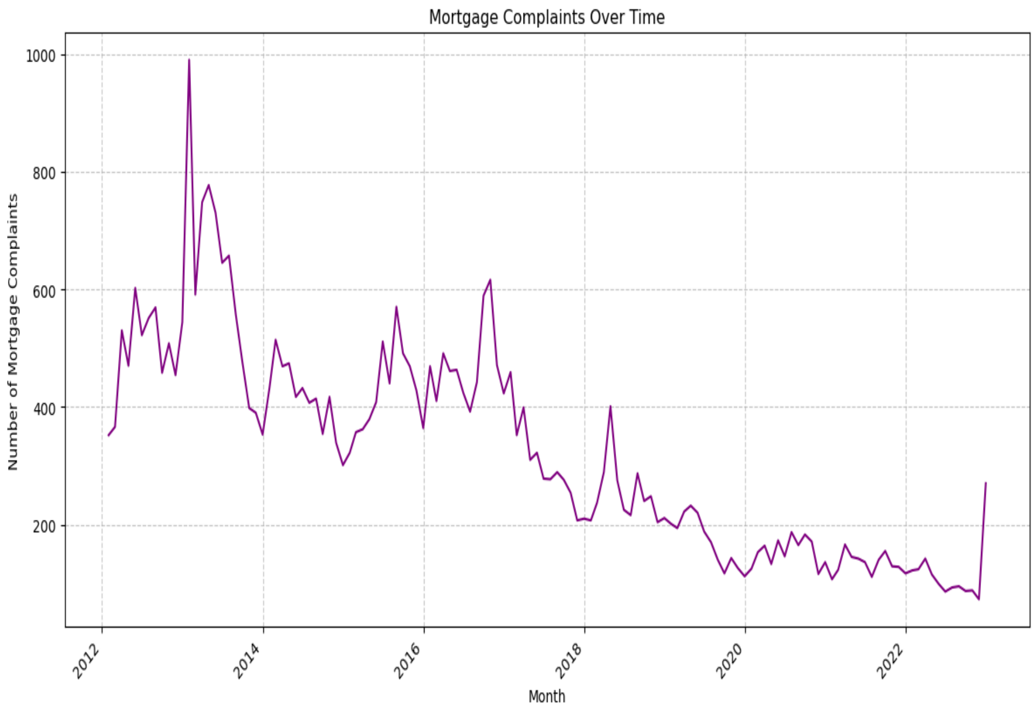

Figure 6. Mortgage Complaints Over Time

From January 2012 to December 2013, the number of complaints exhibits an upward trend. Complaints started at 352 in January 2012 and consistently increased to 748 in March 2013, signaling a significant rise in issues. However, this trend didn't continue in 2014, as the number of complaints started to decline and remained relatively stable.

In 2015, the complaints showed some fluctuations but remained below the levels observed in 2013. Notably, there was a peak in August 2015, with 570 complaints, which could be a point of concern for Wells Fargo.

From 2016 onwards, there's a mixture of fluctuations and some increasing trends in the number of monthly complaints. The highest peak occurs in September 2016, with 589 complaints, indicating another potential area of concern.

4.10 Random Forest Model Results

The accuracy of the Random Forest model is approximately 86.29%. This metric measures the model's overall correctness in predicting the outcomes. An accuracy of this magnitude signifies that the model is quite adept at making accurate predictions in most cases. To ensure that Random Forest has high precision and accuracy, we tested it against KNN, the Logistic Regression Model, and the Decision Tree Classifier Model comparatively. The Ensemble Voting Classifier, which trains many models.

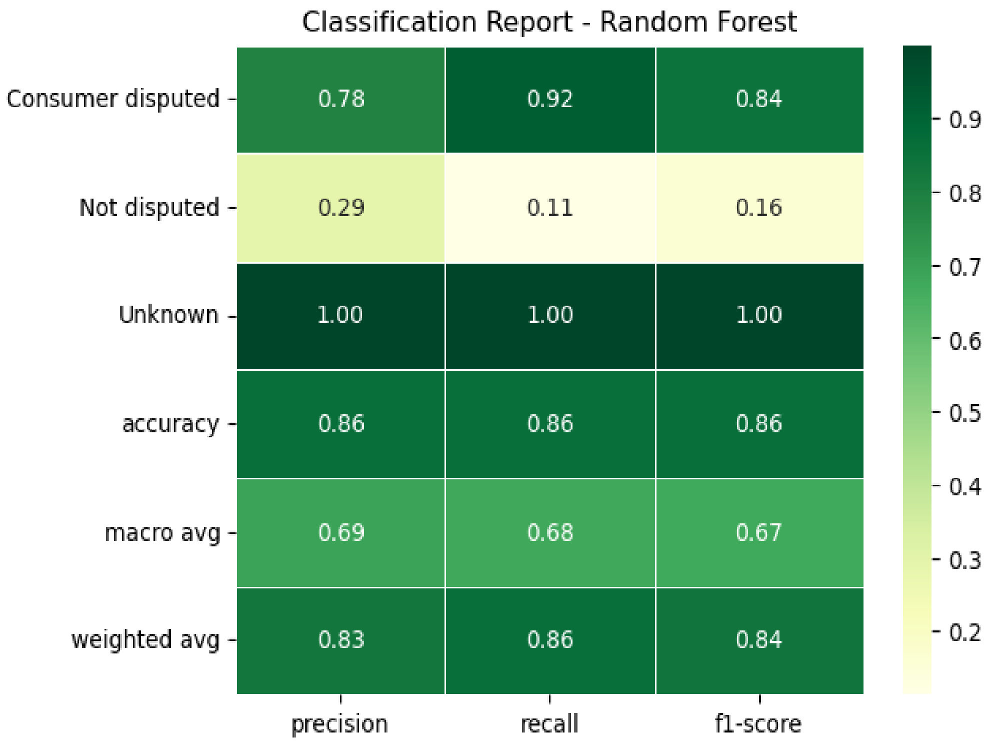

Figure 7. Classification Report for Random Forest

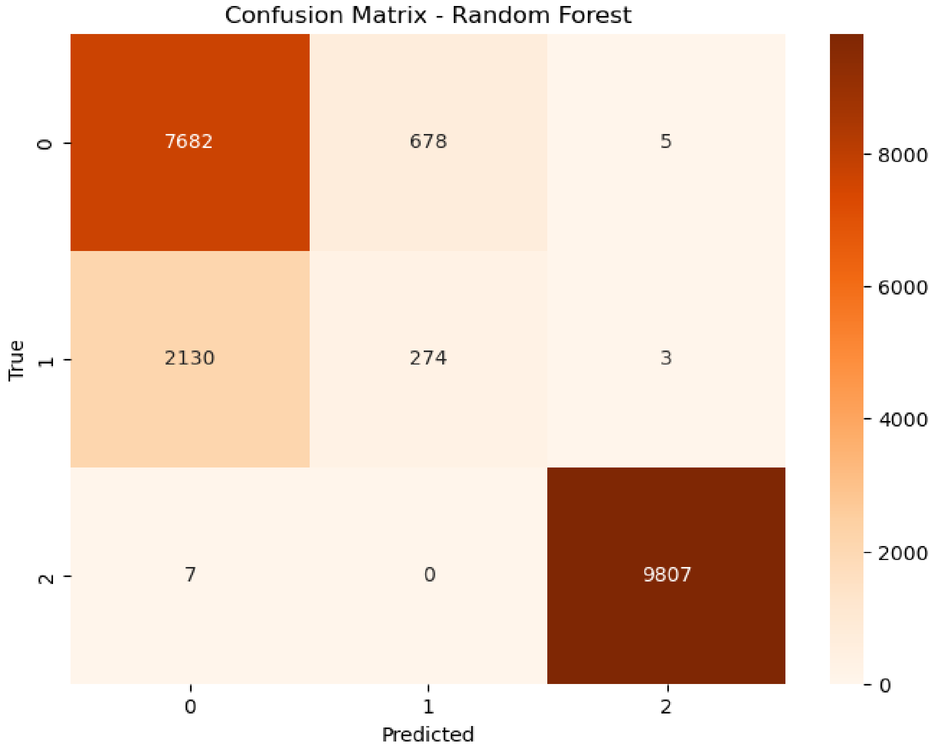

Figure 8. Confusion Matrix for Random Forest

I tested four models and found Random Forest model is having better precision or recall compared to other models that were tested with limited resources and further research is required to improve the model performance.

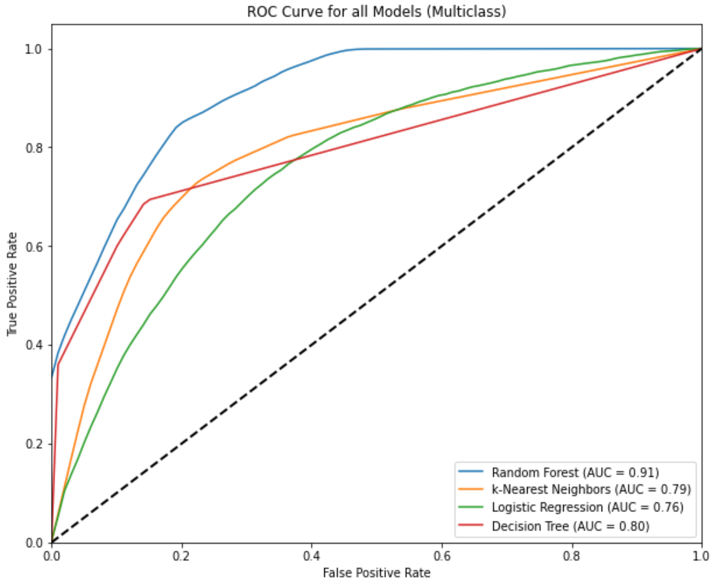

Figure 9. ROC Curve for all the Models

Here we can see clearly that our model accuracy is 0.91 which is way better than other models. The AUC scores collectively reflect the model's effectiveness in meeting the project's primary objective—accurate prediction of consumer complaint outcomes. The AUC scores, measuring the discriminative power of Wells Fargo's consumer complaint model, reveal crucial insights for operational decisions. With a high AUC the model effectively identifies complaints, instilling confidence for Wells Fargo in decision-making regarding non-contentious issues. These scores not only guide resource allocation by focusing efforts on challenging cases.

CHAPTER FIVE

DISCUSSION

5.1 Interpretation of Results

The results obtained from the data analysis and modeling in this research provide significant insights into Wells Fargo's consumer complaints. Understanding these results is crucial in the context of improving customer service, addressing specific issues, and making informed decisions. In this section, I will interpret the findings and their implications for Wells Fargo.

5.2 Temporal Patterns

The analysis of consumer complaints, over the decade provides insights. There has been an increase in complaints from 2012 to 2015 indicating a rise in customer dissatisfaction. During this period there were also some fluctuations that could be attributed to factors like conditions, regulatory changes or shifts in customer sentiment. Notably the highest number of complaints occurred in 2017. This upward trend suggests that Wells Fargo faced challenges in meeting customer expectations during this time. These challenges may have been influenced by high profile scandals and regulatory issues that affected the bank. It is crucial for Wells Fargo to investigate the reasons behind these complaints and take actions to prevent them from happening again.

Although there was a decrease in complaints in 2019 and 2020 it's important to note that there was an increase towards late 2021 and a significant spike in December 2022. This should raise concerns for Wells

Fargo. They need to understand what caused these spikes whether they are related to issues or if they indicate a trend. This requires monitoring and prompt responsiveness, to emerging issues.

5.3 Complaints by Product Type

The analysis of customer complaints, across product categories highlights areas that Wells Fargo should prioritize. It's clear that the highest number of complaints are related to mortgages followed by checking or savings accounts and general banking services. These findings indicate that Wells Fargo needs to address concerns in their mortgage department as in their basic banking services, which have been a significant source of dissatisfaction for customers.

To tackle these issues effectively Wells Fargo should consider implementing an approach. This may involve reviewing and improving processes providing training for customer service representatives and enhancing communication, with customers regarding mortgage terms, fees, and other important details. Similarly, it's crucial to address any issues related to checking or savings accounts by streamlining account management processes minimizing fees where possible and ensuring the security of customer accounts.

5.4 Common Issues

For Wells Fargo to better understand the challenges faced by their customers it is crucial for them to analyze the issues that arise. It is of importance for Wells Fargo to address the concerns, such, as "Loan modification, collection, foreclosure " and "Managing an account." These issues encompass problems related to loan modifications, debt collection and managing accounts – all of which're aspects of Wells Fargo’s services. By addressing these concerns and implementing processes for loan modifications and debt collection as well as enhancing account management with reduced fees and improved security measures Wells Fargo can make a significant impact, on improving customer satisfaction levels while upholding its reputation.

5.5 Complaints Distribution by State

It seems that when it comes to complaints there are states, like California, Florida and Texas that consistently have numbers of them. This indicates that customer concerns are not the same across all states due to differences in Wells Fargo’s operations or the demographics of its customers.

To address this Wells Fargo should tailor its customer service and complaint resolution strategies based on regions. By considering the concerns and needs of customers, in each area the bank can allocate resources effectively. Provide better support to its customers.

5.6 Implications for Wells Fargo Continuous Monitoring

The findings from this research have significant implications for Wells Fargo:

Wells Fargo must implement a system for continuous monitoring of consumer complaints. The spikes in late 2021 and December 2022 indicate that emerging issues can have a significant impact on complaint volumes. Being responsive to these emerging challenges is essential to prevent widespread customer dissatisfaction.

5.7 Product and Service Improvements

The focus should be on improving areas with the highest complaint volumes. Wells Fargo needs to address mortgage-related issues, improve basic banking services, and streamline loan modification, debt collection, and account management processes.

5.8 Regional Tailoring

Recognizing the geographical variation in complaints, Wells Fargo should tailor its customer service strategies to specific regions. Understanding regional customer concerns can lead to more effective resource allocation and localized solutions.

5.9 External Factors Analysis

To explain the temporal patterns in complaints fully, Wells Fargo should consider analyzing external factors such as economic conditions and regulatory changes. Collaborating with experts in economics and finance could provide a more comprehensive understanding of these patterns.

5.10 Customer Relationship Management

The bank should enhance its customer relationship management (CRM) to proactively address issues. Open communication with customers, prompt issue resolution, and a focus on improving customer experiences can reduce complaints.

CHAPTER SIX

CONCLUSION AND FUTURE WORK

6.1 Summary of Findings

In this study, a comprehensive analysis of consumer complaints directed at Wells Fargo was conducted to gain insights into the bank's customer satisfaction and identify areas for improvement. I tested four models and found Random Forest model is having better precision or recall compared to other models that were tested with limited resources and further research is required to improve the model performance.

The research incorporated exploratory data analysis to understand the dataset's characteristics and uncover key trends, as well as the development of predictive models to provide insights into future complaint trends.

6.2 Temporal Patterns

Complaints exhibited a general upward trend from 2012 to 2015, with some seasonal fluctuations. Notable spikes in complaint volumes were observed in late 2021 and December 2022, warranting further investigation.

6.3 Complaints by Product Type

Mortgage-related issues were the most common, indicating a need for attention in this area. Checking or savings account issues and general banking complaints also had significant volumes, highlighting areas for improvement.

6.4 Common Issues

The most common issues included "Loan modification, collection, foreclosure" and "Managing an account," suggesting the need for enhancements in these processes. Addressing payment difficulties, inaccuracies in credit reports, and financial hardships in paying mortgages were also identified as priorities.

6.5 Complaints Distribution by State

Complaint volumes varied by state, with California, Florida, and Texas having the highest numbers.

Regional differences indicate the need for tailored customer service approaches in different states.

6.7 Future Research Directions

While this study has provided valuable insights into consumer complaints and their implications for Wells Fargo, there remain numerous avenues for future research in this domain. Some potential research directions that could further enhance our understanding and help the company refine its customer service strategies:

6.8 NLP and Sentiment Analysis

Future research can employ advanced Natural Language Processing techniques to conduct sentiment analysis on the text descriptions of complaints. This would enable Wells Fargo to gain a deeper understanding of customer emotions and tailor responses accordingly.

6.9 Real-Time Monitoring

Developing real-time complaint monitoring systems can be instrumental in identifying emerging issues promptly. Such systems could use machine learning algorithms to spot trends and anomalies in customer complaints as they occur, allowing for rapid responses.

6.10 Customer Feedback Integration

Wells Fargo can explore methods to systematically integrate customer feedback, both positive and negative, into its complaint resolution and service improvement processes. This would foster a more customer-centric approach.

6.11 Compliance and Regulatory Research

As regulatory environments evolve, research on the impact of new regulations and compliance standards on complaint volumes and types can be invaluable for ensuring Wells Fargo's adherence to legal requirements.

6.12 Customer Segmentation

Investigating customer segmentation based on complaint types and characteristics can provide a nuanced understanding of diverse customer needs. Wells Fargo can then tailor its services more precisely.

6.13 Cultural and Regional Factors

Research exploring the influence of cultural and regional factors on customer complaints can aid in creating strategies that resonate with specific customer demographics in various geographic locations.

REFERENCES

Altuntas, M., Asutay, M., & Holt, D. (2019). The Impact of Customer Complaints on Bank Reputation: The Case of Turkey. International Journal of Business and Management, 14(9), 97-113.

Chen, J., Cao, Y., & Zhu, L. (2021). Enhancing Customer Complaint Management Through Big Data Analytics: Evidence from the Banking Industry. Frontiers in Artificial Intelligence, 4, 46.

Consumer Financial Protection Bureau. (2013). CFPB Finalizes Rule to Improve Information About Access to Banking Services.

Consumer Financial Protection Bureau. (2020). Consumer Complaint Database: Mortgage.

Deloitte. (2021). AI-Powered Chatbots Transforming Customer Service. Deloitte.

Fay, J., Stone, P., & Zhang, L. (2020). Consumer Complaints and Consumer Harm in Banking Services: Evidence from Bank Ombudsman Cases. Journal of Banking & Finance, 115, 105803.

Gaurav, K., Kaur, P., & Aggarwal, G. (2021). Banking Service Complaint Resolution: A Study of Indian Private Sector Banks. Journal of Business Research, 131, 456-467.

He, W., Zha, S., & Li, L. (2017). Customer Complaint Management in Banks: Exploring the Effects of Complaint Sources and Service Recovery Measures. Journal of Financial Services Marketing, 22(2), 77-88.

Hendricks, D. (2019). Bank Complaint Resolution: A Comprehensive Review. Journal of Banking & Finance, 107, 105606.

McKinsey & Company. (2021). Personalization at Scale. McKinsey & Company.

Microsoft. (2021). Micro Microsoft. (2021). Microsoft AI chatbot lowers response time by 25%, resolves more than 90% of customer issues. soft AI chatbot lowers response time by 25%, resolves more than 90% of customer issues. Microsoft.

Sharma, R., Saini, P., & Agarwal, V. (2020). Predictive Analytics in Banking for Customer Complaint Resolution. Procedia Computer Science, 167, 1933-1943.

Suki, N. M. (2016). Online Customer Complaints: Implications for Digital Marketing. Procedia Economics and Finance, 37, 270-277.

Verhoef, P. C., & Lemon, K. N. (2020). Successful Customer Complaint Management in Multichannel Settings. Journal of Service Research, 23(2), 124-140.

Xu, X., Song, X., & Li, X. (2020). Intelligent Customer Complaint Resolution in Banking: A Chatbot-Based Approach. Future Generation Computer Systems, 110, 635-643.

Zhu, F., Dong, X., & Zheng, J. (2018). Natural Language Processing in Customer Complaints Management: A Review and Future Research Directions. Expert Systems with Applications, 96, 42-55.