Analyzing Consumer Complaints from Financial Firms

Aryan Vakharia
September 30th, 2022

Introduction:

It is indisputable that as our financial institution enters alpha, consumer complaints disputing the financial products offered are definitely going to be a major part of customer service in terms of time and cost. Thus, the following visualizations and insights will help in understanding the factors that influence the disputes in an efficient, accurate manner.

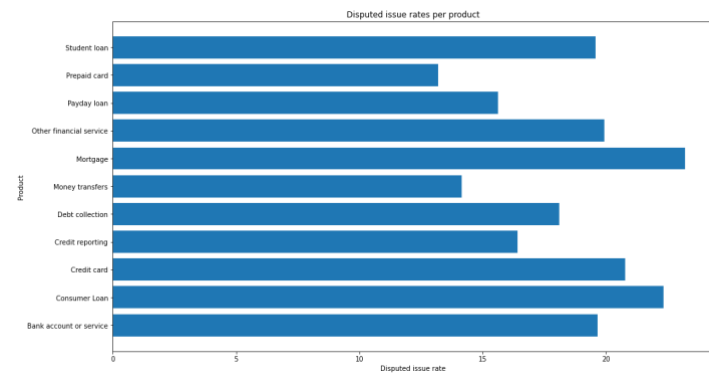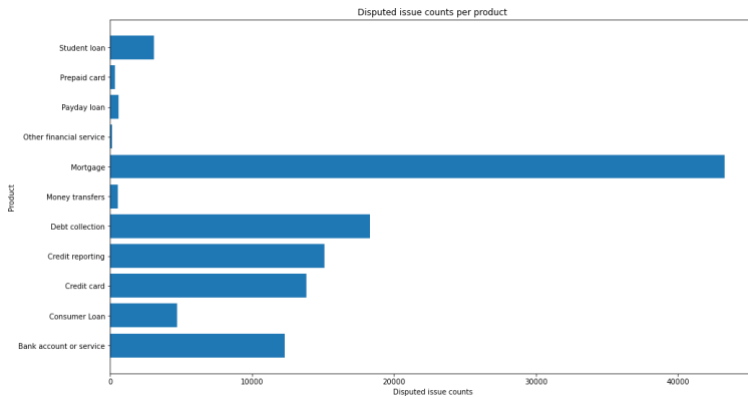## Insights:

Distribution of Disputed Issues over States:
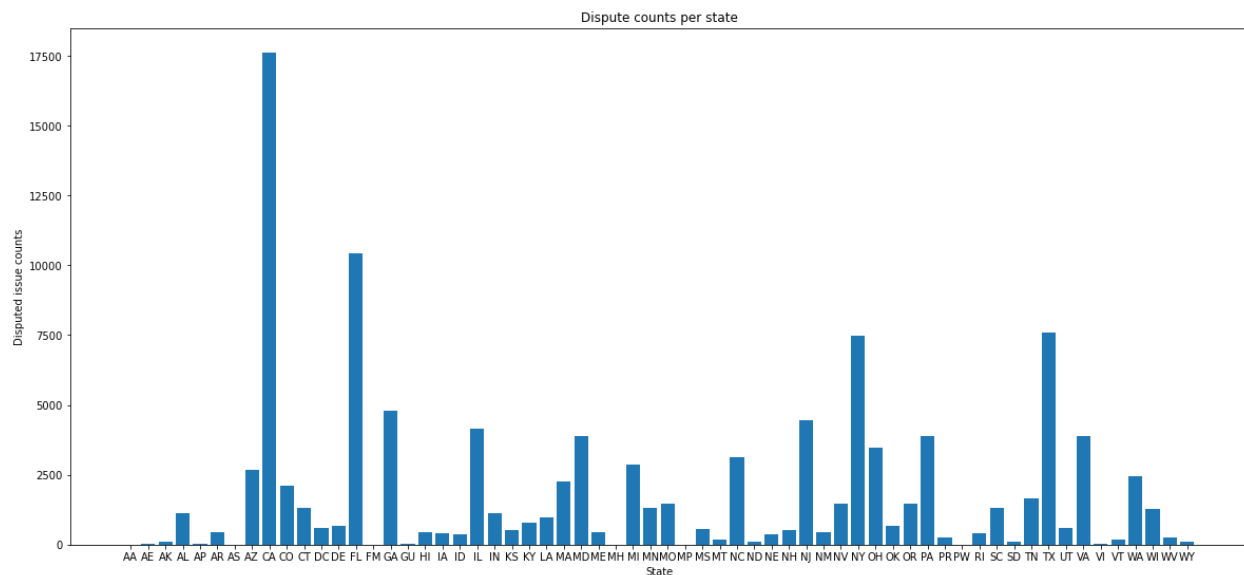
As the bar graph above highlights, the state of California has the most disputed claims(~17500). This is representative of the trend that states in the West Coast (closely followed by those in the East Coast) have the most disputes among consumers.

This insight yields the following guiding questions that can be used for further analysis:
- *What regulations/laws do the states with higher disputes have?*
- *What is the background of the financial institutions based in the high dispute count states?*

*Distribution of Disputed Issues and Disputed Rates over Products:*

As this bar graph shows, Mortgages are the most frequently disputed claims, however it must be noted that while Credit reporting has a lower count, it has a high rate of dispute.

This insight yields the following guiding questions that can be used for further analysis:
- *What causes mortgages to be disputed so frequently?*
- *What causes credit reporting to be disputed so frequently?*

Distribution of Disputed Issues and Disputed Rates over Submission Methods:
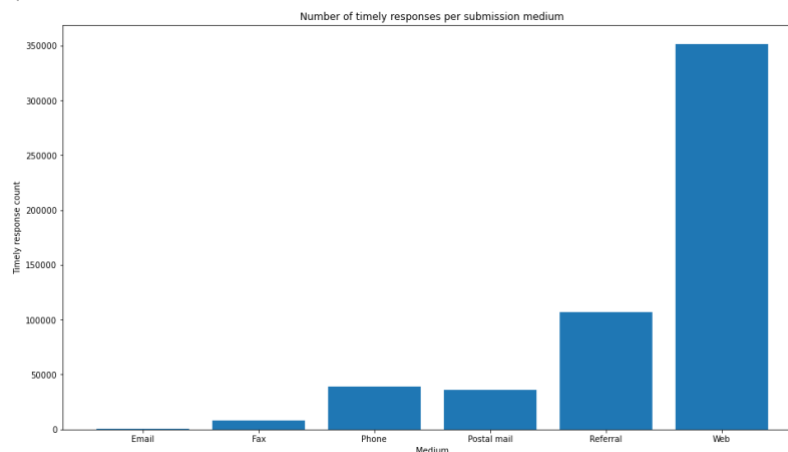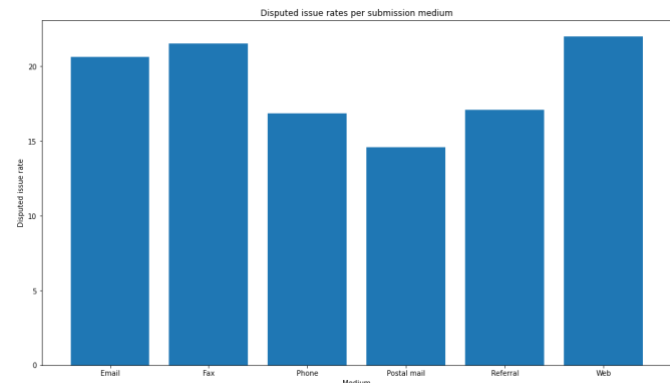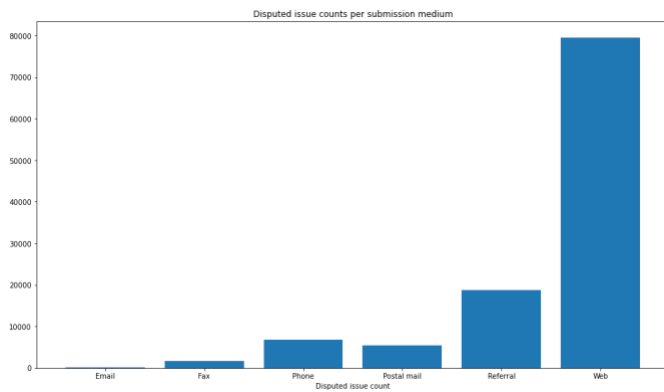
As the two bar graphs above show, digital forms of communication such as Email, Web, etc garner significantly higher counts and rates of disputes as compared to non-digital forms like Referral, Postal mail etc. Another key insight is that non digital communication yields the slowest response times as it may reflect company bureaucracy and general complacency in handling all types of possible communication.

This insight yields the following guiding questions that can be used for further analysis:
- *Can non digital forms be eliminated completely without hindering company operations?*
- *How can digital forms of communication be optimized?*

Results of Dispute Prediction:
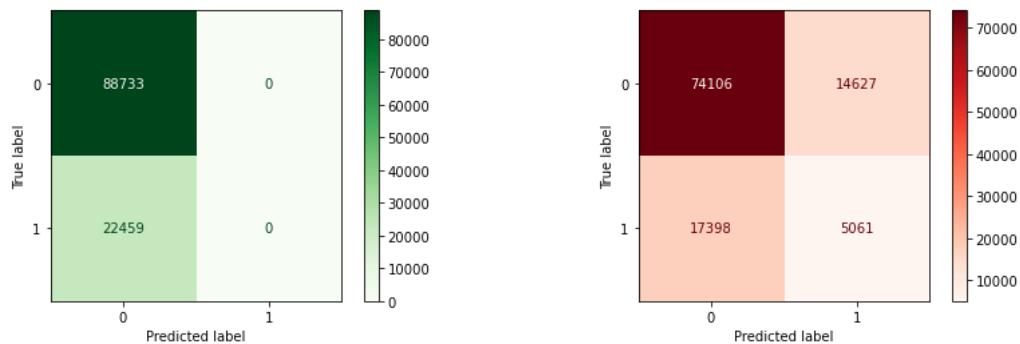
(Green = Logistic Regression, Red = Decision Tree)

While both are not accurate, both classifiers have an accuracy more than 70%. However, with packages like XGBoost and NLTK as well as proper conversion of the issue data, it is possible to predict disputes over issues with higher accuracy.

Conclusion:
Issues pertaining to loans tend to have the most disputes, so if our institution is wary of such issues and can use a fine-tuned version of the above models to predict and forecast the possibility of such issues the operations of the same can become more efficient and well-prepared over the long term.