Financial Complaints Project Future Hypotheses

Anuj Singla

September 29th, 2022

Chapter 1

Current Project

1.1 Optimizing the ML Model

Only one scikit-learn Logistic Regression model was used and was biased toward classifying a complaint in the "non-dispute" category.

1.1.1 If an ensemble method such as XGBoost is optimized for the data set, then the model will make less biased predictions than the Logistic Regression model.

1.1.2 If the data is balanced to include less non-disputes, then the ML classifier will train with less bias.

1.1.3 If stronger feature analysis and engineering is performed on the data set, then the ML model will have greater performance with respect to consumer disputes.

1.2 Data Insights

The following features were compared to consumer disputes: product, state, submission method, and issue.

1.2.1 If a product is more commonly used within a financial institution, then it will have more complaints.

1.2.2 If a consumer is located in a state on the western coast of the United States, then they are more likely to dispute their resolution.

1.2.3 If a consumer uses an online method of submitting their complaint, then they are more likely to dispute their resolution.

1.2.4 If a consumer submits an issue for servicing payments, then they are more likely to dispute their resolution.

1.3 Additional Dispute Prediction Project Implementations

1.3.1 If sentiment analysis on issues is used as a feature in the ML Model, then the classifier will have stronger performance.

1.3.2 If dimensionality reduction with principal component analysis (PCA) is used, then strong, new visualizations and insights can be drawn from the data.

Chapter 2

Additional Ideas

2.1 Data Exploration

2.1.1 Interactive map that highlights different zip codes and their attributes.

2.1.2 Advanced sentiment analysis on product, sub-product, issue, and sub-issue.

2.2 ML Prediction

2.2.1 Predicting how the company will resolve the complaint