

Classifying Loan Consumer Complaint Data

When it's time to take out a mortgage on a house, buy a car, or to cover short-term expenses, there is a wide variety of companies from which one can take out loans from. At the same time, data analysis and advanced modeling have made applying and receiving loans more automated and streamlined than ever before. However, with these highly automated and complex systems has come a litany of consumer complaints ranging from issues with payment to problems dealing with loan modifications and foreclosures. This presents a unique challenge for the consumer faced with an ever-widening market for loans: **how can consumers accurately assess loan companies and the problems they may face when taking out a loan from them? How can consumers protect themselves from the challenges that many borrowers face when issues arise with their loans?** To this end, our team has applied natural language processing methods to build a model that attempts to categorize complaints by issue, with the goal of painting a clearer picture of what issues different customers face when borrowing from various companies.

Following the financial crisis of 2008, the Consumer Financial Protection Bureau was established as authorized by the Dodd-Frank Act. The CFPB has a wide range of oversight responsibilities, but one of its key roles is to collect and track consumer complaints about various companies and their loan processes. This data set of consumer complaints from the CFPB comes with a categorization of the nature of the issue that the consumer is facing. Because of this pre-existing categorization, this data set is perfectly positioned for building a supervised learning NLP classification model. With this model in hand, **new consumer complaints from outside the CFPB could be categorized by issue and provide consumers a look into the problems that customers of any given company face and how often.**

Beyond the consumer, lenders have a vested interest in being able to accurately categorize consumer complaints to understand where they have room to improve their services. While data from the CFPB comes with pre-established categories that have been manually labeled, **many companies may be sitting on decades of consumer complaints with no big-picture perspective on the most prevalent issues that plague their customers.** Our project can categorize these large complaint datasets and provide direction for improving the customer experience.

Our models were able to perform better than random sampling at categorizing mortgage-related complaints from the CFPB data set, achieving a micro F1-score of .69 using the RoBERTa model for natural-language processing. That said, there may be more avenues in the future to explore different categorization options and hyperparameter tuning. The eventual goal of this model is to extend it to apply to all different types of loans and allow consumers to explore what categories of problems the customers of any given company face in order to make a more informed decision about lending. **Taking out mortgages or student loans are often the biggest financial decisions that individuals make in the course of their lives, and this model can be an invaluable tool in assessing the ever-widening market in lending.**