Predicting the Outcome of Consumer Finance Complaints

Millie Dwyer, Zahra Kadkhodaie, and Zhanna Zhanabekova

New York University, May 2017

Background

- In 2011, the Consumer Financial Protection Bureau (CFPB) was established in response to the Financial Crisis of 2007-2008 and Great Recession
- The CFPB is an independent federal agency responsible for consumer protection in the financial sector
- It has handled over 1 million complaints, helping consumers connect with financial companies to get direct responses about problems with mortgages, student loans, payday loans, debt collection, credit reports, and other financial products and services
- https://www.consumerfinance.gov

Project Objective

- Predict the type of resolution offered for a consumer finance complaint, particularly whether the complaint is closed with monetary relief, using the CFPB 2013-2017 data

Why This is Important

The analysis may:

- Help customers form better complaints
- Give companies insights about the sources and patterns of complaints, and potentially use of the predictive model to automate the process of resolution
- Help the CFPB spot potential violations in the financial marketplace

Data Description

- The data is publicly available from the CFPB
- We detected inconsistencies in labeling in 2011-2012, so we used data from 2013 to 2017. N $\approx$ 660K
- The data contains categorical features such as product (e.g., bank account), issue (e.g., billing statement, loan servicing), date of complaint submission, complainant's zip code and state, etc.
- We chose Company response to consumer as the outcome variable. It originally contained seven categories including 'Closed', 'Closed with explanation', 'Closed with monetary relief', 'Closed with non-monetary relief', 'Closed with relief', 'Closed without relief', 'In progress' and 'Untimely response'
- About 20% of all complaints have actual complaint narratives
- We also added data on zip code-level population characteristics from the American Community Survey (e.g., age, race, income)

Challenges and Feature Engineering

- Initially the multiclass outcome was transformed to a binary variable by merging all classes except the monetary relief. We dropped 'In progress' class.
- This created severe imbalance in the binary outcome (7% vs 93%) as well as ambiguity and fuzziness in the outcome classes. To address these issues, we restricted modeling to a subset of data which contained only complaints labeled as 'Closed with monetary relief' and 'Closed with non-monetary relief'
- Furthermore, because monetary relief is most typical for the top three products - credit cards, mortgages, and bank accounts, - we created another subset of data containing only these top three products
- Weighing outcome classes in the original dataset did not improve performance significantly. So we narrowed down the scope of analysis to the restricted dataset
- Also, we tried splitting data into train, validation and test sets based on the time order of complaints instead of randomly splitting the data. Performance did not improve significantly

Summary of Results (Using Validation Data)

Model	Accuracy	AUC	Precision	Recall
Majority Rule	0.623	-	-	-
Classification Tree	0.743	0.804	0.779	0.820
Random Forest	0.734	0.790	0.756	0.847
Adaboost	0.729	-	0.756	0.834
Logistic Regression	0.815	0.897	0.714	0.754

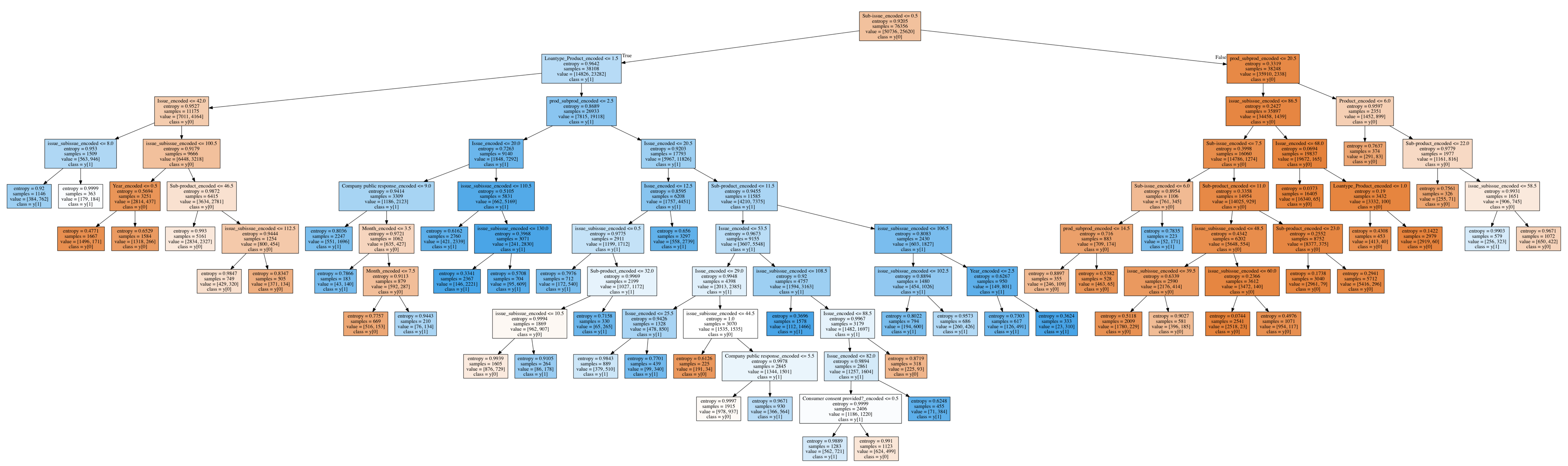

Figure: Regularized tree

Methods

We experimented with various feature sets as well as different models. The present results are derived from analysis on the best feature set

Best Feature Set:

- Restricted feature set with top three products and binary outcome produced by dropping all classes except monetary and non-monetary resolutions

Performance Evaluation:

- Because of the unbalanced nature of the outcome, we rely on several metrics, including accuracy, AUC, precision and recall, to evaluate the performance of each model

Algorithms:

- Baseline: Predicting the dominant class of the training set (no monetary relief)
- Regularized Classification Tree
- Random Forest
- Adaboost with Decision Tree
- Regularized Logistic Regression

Results

- As shown in the table, there is no single model that outperforms all the other models by every evaluation metric. Logistic regression has the highest AUC of ROC and accuracy, while Adaboost and decision tree have the highest precision and recall
- Given the imbalanced nature of the outcome classes, in our final decision, we place less emphasis on accuracy
- From the perspective of a financial company, predictions with low precision and low recall might be equally costly or misleading. So, we choose the best model based on a combination of precision and recall or the F-score. Accordingly, classification tree is our model of choice for analyzing this data and predicting whether a complaint would be resolved with monetary relief

Conclusion

- In order to predict the type of resolution, monetary vs. non-monetary relief, offered by financial institution we performed multiple predictive analysis, using tree-based algorithms as well as logistic regression
- We did feature engineering in multiple rounds. Initially we used all the complaints available, with outcome variable produced by merging relevant categories. Due to inconsistent labeling in the outcome data, we decided to re-create outcome variable by dropping ambiguous categories and keep only complaints with explicitly monetary and non-monetary labels. Additionally, we constrained the dataset to only the top 3 financial products. By focusing only on the financial products with frequent monetary relief outcomes and being more discriminate in the choice of labels, our modeling resulted in better performance along accuracy, recall, as well as precision

Potential Future Work

- Almost 20% of the complaints have complaint narratives. We included in out data the presence of such narratives as a binary variable. Analyzing these text data using natural language processing methods may add to the predictive power of model
- One feature in the original dataset was Company with more than 4,000 unique company names. Because of the large number of categories, this variable was not included for analysis. By mapping individual companies to larger business structure categories, one can reduce the number of categories and create a potentially powerful predictor