Predicting Bond Amounts in Orleans Parish Criminal District Court

TULANE UNIVERSITY
SCHOOL of SCIENCE & ENGINEERING

Advisor: Dr. Aron Culotta
Will Rodman: BS Mathematics, Computer Science
Bharat Solanky: BSM Finance, Computer Science

ABSTRACT

Our project explores the potential of machine learning to predict bail amounts within Orleans parish court records and understand the demographic factors that influence machine learning models. Working in collaboration with Court Watch NOLA and guided by Dr. Aron Culotta, we utilize over 800,000 court records from the Orleans Parish District Criminal Court, focusing on 9,800 cases between 2020-2022 to build a model capable of estimating initial bond values. Inputs for our model include charge type, magistrate, and defendant demographics, among others, clustered into 40 categories for analysis. Preliminary results, achieved through comparing several regression models, show that Support Vector Regression outperforms others with a significantly lower mean percentage error, suggesting machine learning's efficacy in estimating bond amounts accurately. Further analysis of the model's features indicated a negative correlation between bond amounts and the defendant's age, answering critical research questions and highlighting areas for future exploration on bias and fairness within bail determinations.

CHARGE FEATURES

The number of charge descriptions and IDs in the Court Watch NOLA database needed to be reduced to produce a feature matrix with adequate dimensions for a regression model. This was accomplished by grouping charge IDs with similar descriptions into a new, generalized charge ID. The following illustration serves as a hypothetical illustration.

Charge ID	Charge Description
2	DRIVER'S LICENSE
4	VALID LICENSE PLATE

→

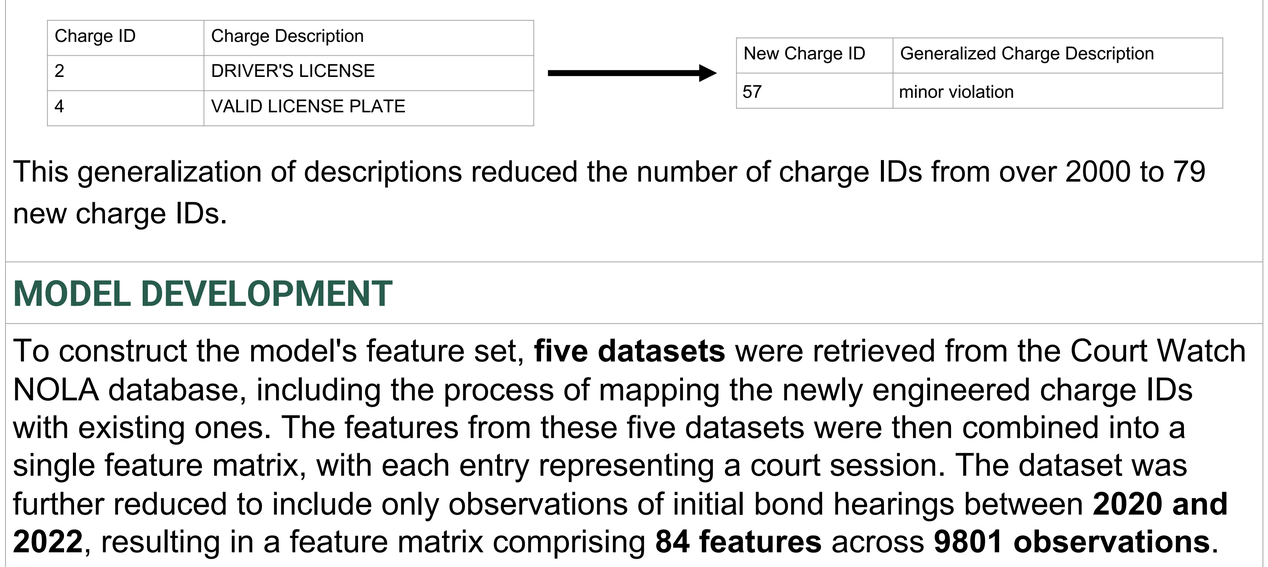
New Charge ID	Generalized Charge Description
57	minor violation

This generalization of descriptions reduced the number of charge IDs from over 2000 to 79 new charge IDs.

MODEL DEVELOPMENT

To construct the model's feature set, **five datasets** were retrieved from the Court Watch NOLA database, including the process of mapping the newly engineered charge IDs with existing ones. The features from these five datasets were then combined into a single feature matrix, with each entry representing a court session. The dataset was further reduced to include only observations of initial bond hearings between **2020 and 2022**, resulting in a feature matrix comprising **84 features** across **9801 observations**. The table below presents the five most highly correlated and the five least correlated features within this feature matrix, in relation to bond values. The following histogram illustrates the distribution and skew of bond values.

Feature Name	Correlation to Bond
accessory to murder	0.43
intimidation/obstruction	0.202
violent	0.119
major case of burglary w firearm	0.086
major case case w weapon (not firearm)	0.083
domestic abuse	-0.05
property.crime	-0.052
age	-0.056
domestic.violence	-0.087
stay_away_order	-0.095



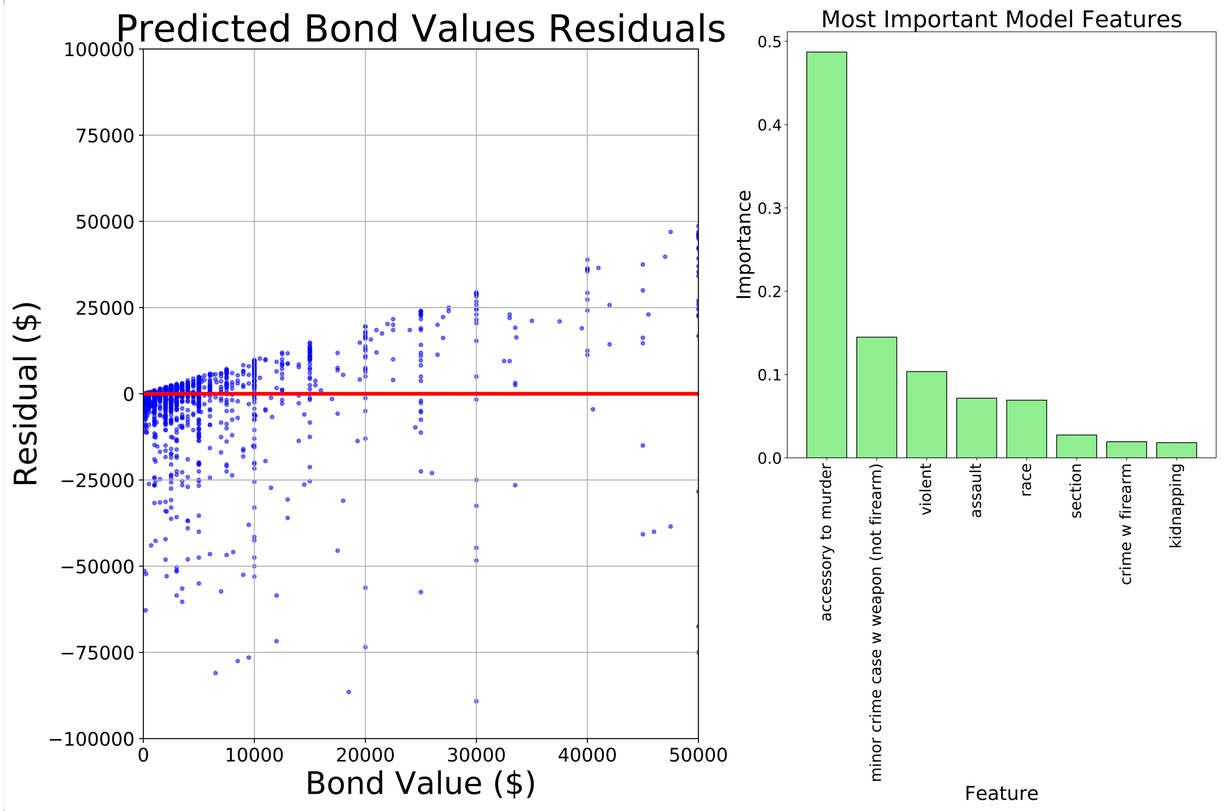
MODEL RESULTS

For the model pipeline, the feature matrix was divided into separate datasets for training (70%), testing (20%), and validation (10%). Initially, four regression models were evaluated using the validation dataset, resulting in the following mean absolute percentage errors (MAPE).

$$\mathrm{MAPE}(y, \hat{y}) = \frac{100\%}{N} \sum_{i=0}^{N-1} \frac{|y_i - \hat{y}_i|}{|y_i|}.$$

Model	Support Vector	Random Forest	Decision Tree	Gradient Boosting
MAPE (%)	3.45	4.94	4.95	7.70

Selecting Decision Tree Regression, through parameter optimization and tested on the training dataset, its **MAPE was reduced to 4.61%.** The residuals from this test are depicted in the scatter plot, with **mean and median residuals of -398 and -608**, respectively, indicating the model's tendency to overestimate values. The bar chart displays the most significant features for the model, highlighting the disproportionate importance of **a defendant's accessory to murder** in predicting bond values.



Calculating the error rates by the race of defendants reveals a difference in the distribution of errors, as shown in the following table.

Race	Sample Size	MAPE (%)	Mean Error ($)	Mean Bond Value ($)
Black	1536	3.97	843.37	22236.33
White	356	7.32	-5929.39	11061.82

Future Work

Possible improvements to the model include providing an ordinal categorical output instead of a regression. This output has potential because magistrates typically set bonds in incremental values. A second improvement could be to include more information about the defendant's court record history in a defendant observation. Finally, it could be beneficial to further test the Support Vector Regression Model, since it received the lowest MAPE.

Thank you to the Court Wach NOLA organization for partnering with the Tulane University computer science capstone program. Learn more about Court Watch NOLA at www.courtwatchnola.org.