Final Results

DS 3001

Anna Girerd, Marisa Guajardo, Owen Himmel,
Clarissa Kusmana, Christine Lee

Colab link: Final_Project_DS3001.ipynb

For our project, we looked at federal defendant charges and sentences over the fiscal year 2018-2019 in the United States. The data set is sourced from the U.S. Sentencing Commission, which collects data on federal sentencing trends annually from all federal District Courts across the country. We decided to narrow our focus to only defendants who had been convicted of trafficking either powder or crack cocaine, and we filtered the data to only include defendants with Criminal History Level 6 so as to control for the influence of repeated offenses. We wanted to analyze the data to investigate: **Which factors are most predictive of sentence length for federal cocaine trafficking offenders, and how plea deals might impact these predictions?** These results and modeling approach could, in turn, be utilized by the Department of Justice or for identifying bias in how sentence length may be determined.

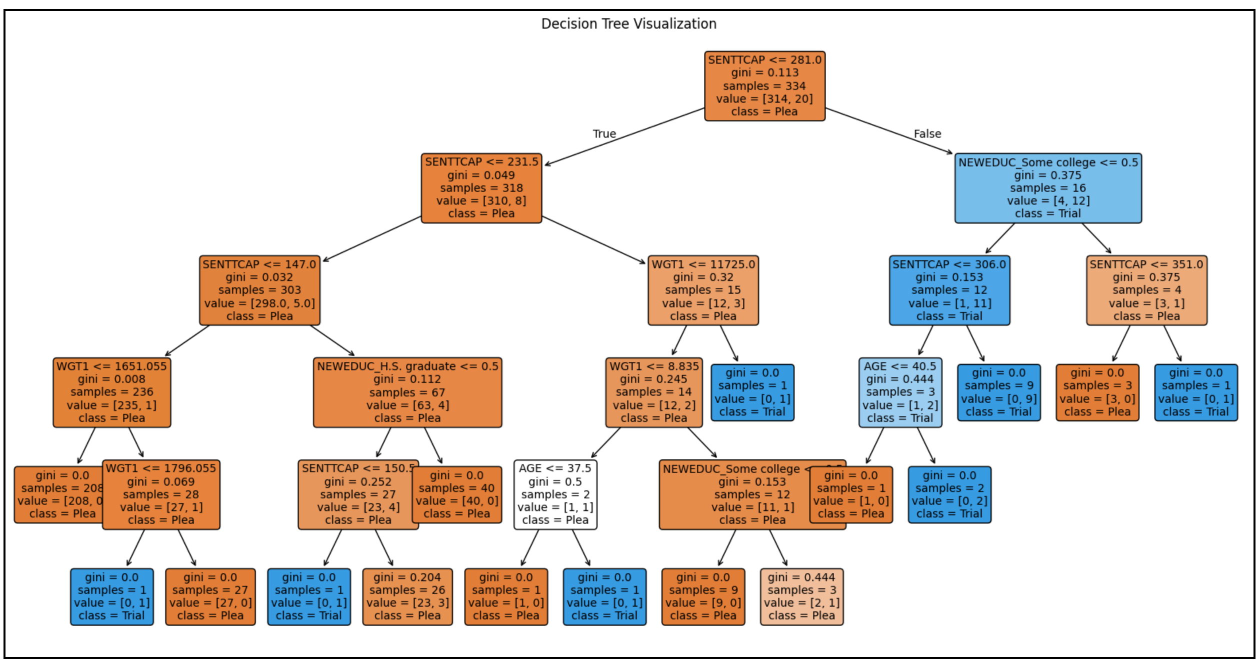

Figure 1: Variable Weights in Plea vs Trial Outcome

The dataset we used was very large and included hundreds of variables. For this reason, we focused on identifying the variables of most importance to our analysis. After cleaning the data, we created a decision tree classifier object with a maximum depth of 5 to prevent overfitting. This decision tree (Figure 1) predicts whether a case is classified as "Plea" or "Trial" based on key features, such as sentence length (SENTTCAP), drug weight (WGT1), and education level (NEWEDUC) of the defendant. From the beginning, SENTTCAP appears to be the most influential factor, with the first split occurring at SENTTCAP <= 281.0. Cases with sentence lengths below the 281-month threshold are predominantly classified as "Plea," while cases with longer sentences move to the right for further evaluation.

As the tree branches out, additional splits refine the classification. For instance, shorter sentences (where SENTTCAP <= 147.0) almost exclusively fall into the "Plea" category, with a high degree of confidence, indicated by the Gini index close to 0. Meanwhile, cases with longer sentences are more likely classified as "Trial," especially combined with factors such as higher education levels or higher drug quantity. Education level also plays a notable role, with defendants having "Some college" are less often classified as "Plea," while higher education levels increase the likelihood of getting a "Trial" classification. Similarly, drug quantity appears in several splits, which means it plays an important role in distinguishing between "Plea" and "Trial" cases in borderline scenarios.

Finally, the tree shows that shorter sentences and lower education levels strongly correlate with the "Plea" group. On the other hand, longer sentences and higher education levels are more correlated with "Trial" observations. The model demonstrates high confidence in many of its classifications, with several leaf nodes achieving perfect purity, where the Gini index is almost 0. However, there does exist an imbalance in class distribution, with "Plea" as the majority, and over 96% of defendants in this population within the dataset taking a plea deal and only 4% taking the case to trial. This imbalance is also present in the structure of the tree. Specifically, at the root node, the value is [314, 20], which means 314 cases are classified as "Plea" and only 20 as "Trial."

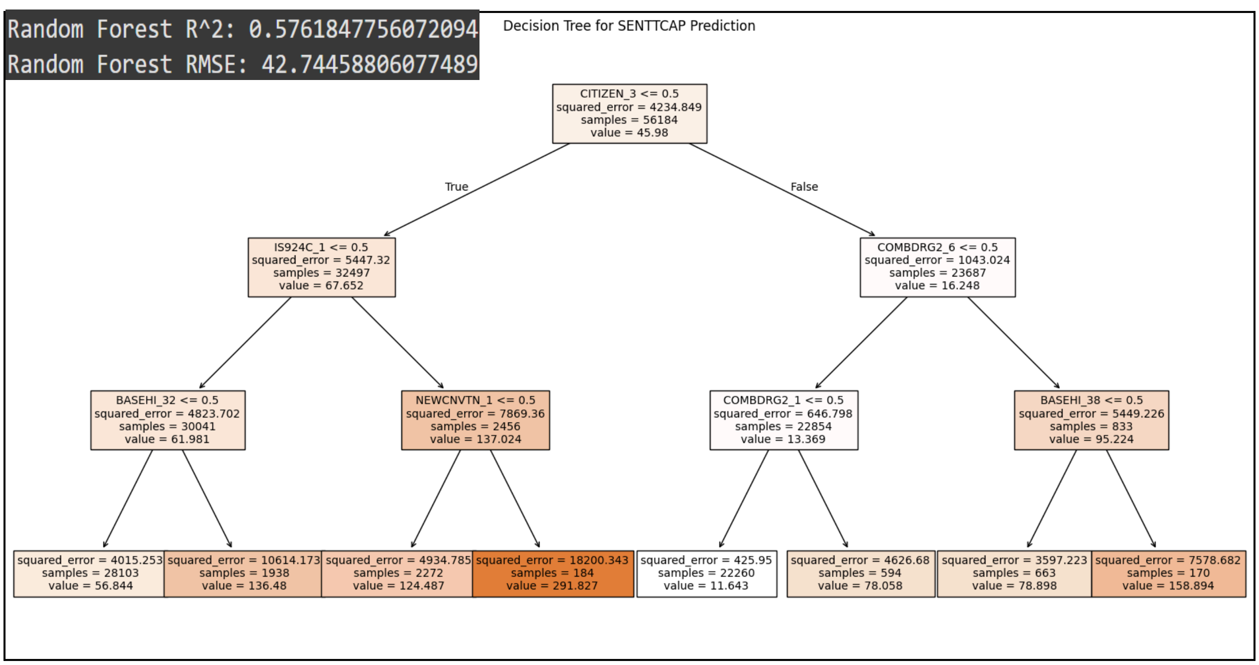

Figure 2: Predicting Sentence Length with Random Forest

Figure 2 shows a sample tree and accuracy statistics for a random forest model predicting sentence length. Variables of interest were citizenship, race, age, education, drug type, and weapon enhancement.The sample decision tree split at illegal alien status, and was followed by splits at weapon enhancement and drug type. Age, race, and education did not appear in the sample tree. Using an 80/20 train test split the model had little efficacy. The RMSE metric showed an average of a ~42 month difference between predicted sentences and real sentences. This suggests a strong omitted variable bias. Because of the complexities of criminal sentencing predictions would need to rely on either more inputs or more nuanced variable choice. Regarding ability for examining potential discrimination, a data set would need to be constructed with near identical crimes. Demographic data appears to not hold significant importance when compared to things like weapon enhancement or drug type. Given the inaccuracy of the model we decided to further refine our predictive strategy with the addition of Random Forest and LASSO Regression modeling as outlined below.

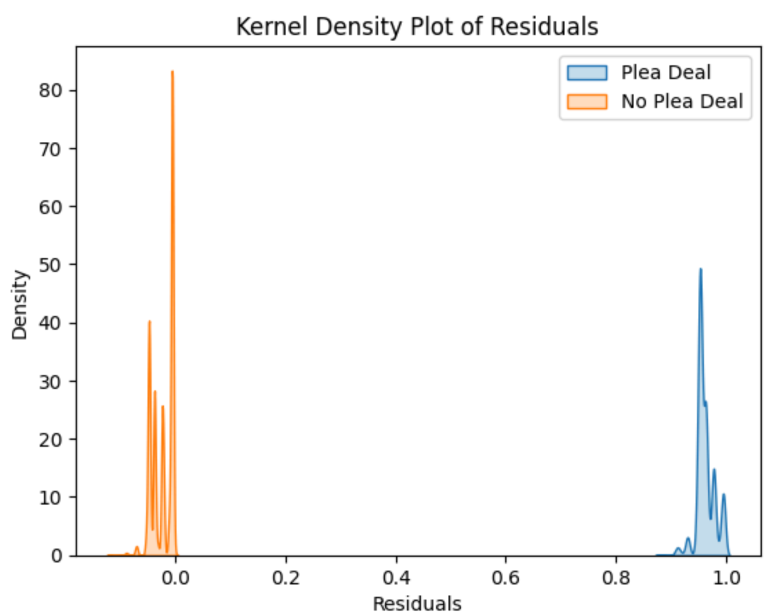

Figure 3: Nuisance Variable Influence on Plea vs Trial Outcome

Next, as demonstrated in Figure 3, we split the data into two features, nuisance variables (X_nuisance) and the target variable (y_target). We used this to train a Random Forest Regressor to predict the target variable using only the nuisance variables. The maximum depth parameter is set to 5 to control the complexity of the model. The model's predictions are used to calculate residuals, which represent the difference between the actual target values and the predicted values. We created a kernel density plot (Figure 3) to visually compare the distribution of residuals from the Random Forest Regressor for two groups: cases with a plea deal and cases without a plea deal. This helps assess whether the nuisance variables have a different impact on the prediction of plea deals for these two groups.

The sharp separation indicates that the two groups are fundamentally different, and nuisance variables alone cannot account for the distinction. This suggests a strong relationship between plea deals and other unmodeled factors and demonstrates that plea deals and no plea deals differ significantly, and the differences are not explained by nuisance variables. The model explains only 2% of the variation in plea deals using dummy-encoded nuisance variables with an R-squared of 0.02. The dataset is heavily imbalanced with 22,230 No Plea Deals versus 589 Plea Deals, which skews the model's performance innately. We would need to address the class imbalance to improve predictions for plea deals.

The influence of whether a plea or trial is taken evidently is large on sentence length outcome. For this reason, we decided to perform a LASSO regression analysis to look at a larger number of variables and their influence. The resulting visualization is quite large, so we've cropped the image to include the important variables.

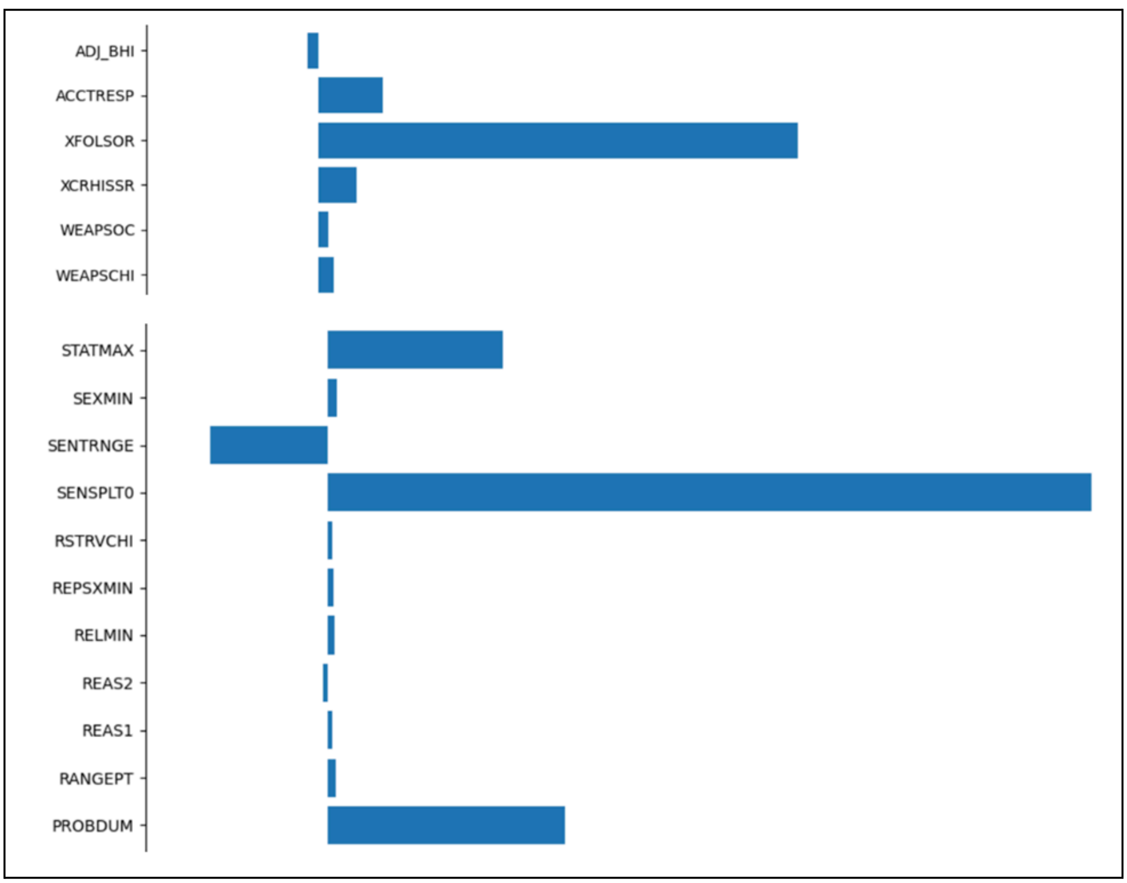

Figure 4: LASSO Variable Weight in Sentence Length

The Lasso reduced the feature set from 211 to 86, making the model simpler and more interpretable without significantly compromising performance. SENSPLT0 was a key feature (Figure 3), but this variable is total prison sentence in months with zeros, which is essentially the same as our response variable, so we disregarded this finding. The variable WEAPSOC (Figure 4) indicates whether there is a Specific Offense Characteristic (SOC) enhancement related to the use or presence of a weapon in a case. If a weapon was involved in the commission of a crime (e.g., carrying, brandishing, or using a weapon), a SOC enhancement could increase the severity of the sentence. While the Lasso indicates that the presence of a weapon enhancement is relevant in the prediction of sentencing outcomes, it surprisingly isn't as salient as other variables. XFOLSOR (Figure 4) seems to have a stronger influence. This is a numeric variable for final offense level as determined by the court on a scale of 1 to 43. PROBDUM is another variable with a relatively stronger influence, and this variable indicates whether the defendant received probation.

We performed a regression analysis on the Lasso non-zero coefficients to see how well the model worked. The R^2 and adjusted R^2 of 0.94 suggest that the model is quite effective in predicting the sentence length (SENTCAP), but the MSE indicates there is still some room for improvement in prediction accuracy.

Through our analysis of this dataset, we have found some interesting takeaways about the impact of different variables on sentencing length for cocaine trafficking. First, demographic variables such as race, sex, and education levels did not have as big of an influence as we first predicted. There are a lot of variables of significance that are related to sentencing guidelines, such as MNTHDEPT, which indicates the difference in months between the guideline minimum and the sentence length. MAND1 indicates the status of any mandatory minimums at sentencing, which was a variable included in the Lasso, but had a smaller influence than other variables like PROBDUM, for example.