Using machine learning to understand the factors shape sentencing outcomes

Chuyuan Zhong[1], Jiaqin Wu[1], and Li Zheng[1]

[1]Georgetown University
[1]This manuscript was compiled on Dec 6th, 2022

Abstract

This paper conducted machine learning approaches to predict sentencing outcomes, specifically, whether to be incarcerated or not and sentence length, and explore the correlation ranking of the variables that participated in the prediction by using the Cook County sentencing data, the Chicago Police Station data set, and the Chicago neighborhood and boundary data. The time range of our predictions includes before and during the COVID-19 pandemic. We found that the prediction of whether being incarcerated or not has a relatively ideal accuracy of 86%. However, the sentence length prediction needs to be further optimized. We believe that it can be further improved by expanding variables and using a larger and complete sample. Therefore, we concluded that, for incarceration prediction, the random forest model proved to be a more efficient and accurate machine learning model. Race, age of the accused at the incident, and type of alleged crime had major impacts on the prediction results.

Keywords: Machine Learning, Chicago Police Station, Incarceration, Sentence Length, Prediction

1 Introduction

Amid a wave of police brutality and the ravaging spread of COVID-19, calls for reform of the criminal justice system seem to be growing louder than ever. Contrary to the advocacy of reduced government intervention, more than two million people in the United States, disproportionately people of color, are incarcerated in prisons, jails, and detention centers [**Sawyer &Wagner**, **2022**]. Thus, much of the debate surrounding the reform is focused on reducing mass incarceration and length of time. In fact, policies implemented in the post-pandemic era achieved some extent of de-incarceration. In Los Angeles, Philadelphia, and Baltimore, the withdrawal of warrants for non-payment of court fines and a reduction in arrests for minor offenses have reduced local incarceration rates [**Macmadu et al.**, **2020**].

Although the current slowdown in the justice system has led to a downward trend in incarceration rates at the local level, the infiltration of the justice system by entrenched institutional racism in

American society has led to numerous injustices, in terms of disparities in sentence length and the size and composition of the incarcerated population. These differences are based on the court's decision which is the result of a complex measurement of combination factors. For example, the crime committed by the defendant, the harm caused to the victim, the criminal impact on society, the historical crime rate in the local area, the court district, and the demographic information of both defendants and judges, etc. Therefore, based on a set of measurable and relevant factors, our research hopes to answer the question of what features are important in influencing the judges' decisions related to incarceration and sentence length in the context of the pandemic.

To explore the factors that drive decisions of being incarcerated and length of sentence, we built different machine learning models using an open sentencing data set provided by Cook County Government, Chicago Police Stations data set and geographic data set of Chicago boundaries and neighborhoods provided by the City of Chicago. Our models took into account the factors of race and gender of defendants, the predicted gender of judges, categories of criminal offenses, age of the defendant at the date of the incident, the number of charges relating to the defendant in a case, neighborhood livability, court districts, and historical incarceration rates as key variables, while also introducing a time indicator to examine how COVID-19 alters court decisions, to predict sentencing outcomes. Our goal is not limited to comparing model performances but includes interpreting how we use machine learning models to predict and the rank of feature importance. Thus, we are able to identify and evaluate what variables are most relevant in contributing to the sentencing outcomes.

2 Related Work

Scholars pointed out that the differences based on the court's decision in sentencing outcomes come from extralegal factors such as race, gender, and socioeconomic status [**Nellis**, **2021**]. Dr. Ashley Nellis, Senior Research Analyst at The Sentencing Project, in her report, mentioned that African Americans are incarcerated in state prisons at about five times the rate of whites, and incarcerated Latinos defendants are at 1.3 times the rate of whites. In felony drug cases, white defendants were more likely to be transferred to drug court than in similar cases against black defendants under similar circumstances where factors were held constant. Other than the chronic racial and ethnic disparity in incarceration, the composition of the neighborhood may also be a factor in the outcome of a sentence, regardless of the race or ethnicity of the accused. Notably, delays in the criminal justice system due to the emergency response to COVID-19 have decreased the prison population. A 25% reduction in the national prison population by 2020, with women's prison populations and incarceration rates declining more than the male population [**Sawyer**, **2022**]. There has also been a marked decline in the rate of juvenile incarceration. Even though some reports suggest that the decline in the prison population is beginning to rebound, the impact of the pandemic has become an important factor in determining sentencing outcomes.

In their research, Ostrom and Kleiman took the data of the sentencing results in Michigan as an example, conducting statistical models to analyze the impact of variables in each category on sentencing outcomes respectively. For example, they picked the gender category to examine the effect of being male or female on the sentencing outcome. By establishing the Multinomial Logit Model and probit model, they were able to enumerate coefficients and statistical significance of different variables. In addition to focusing on variables from a single category, they constructed 50 4x4 sentencing tables to evaluate sentence length using the intersection of crime classes and previously recorded scores. In the end, they conducted a series of Wald tests to use Chi-square statistics to evaluate the relative importance of each regression model. However, most practices of these models focus on the influence of a single category, as the authors explained that the estimated coefficients of multiple logit analyses are difficult to interpret in a meaningful way [**Ostrom et al.**, **2004**].

On the basis of the analytical process of their research, we made changes adapted to our case and further extended them. First, in order to avoid the lengthy and complex category variables when applying them to one model, we extracted variables in every single category as indicator variables with only 1 and 0 values. Then we visualized the relationship between each factor and the predicted results to help us understand the effects of single individual variables before we put them into machine learning models. By doing this, we can see how the factors together contribute to predictive modeling; and by evaluating the feature importance to the model's predictive behavior, we can find out which variables in a specific category are mostly associated with our predictive outcomes.

3 Data

We used three data sets in our research. First, we mainly used the updated version of sentencing data from Cook County Open Data that was released on 28th September 2022 on the Cook County Government website. The Sentencing Data set includes a summary of defendants' prosecution and crime information and sentences, court information as well as basic demographic information. We pulled a lot of data from it including but not limited to offense categories, age of incident, court district, race, gender, and sentence outcomes. Since one defendant can be associated with multiple charges in one case, we focused on the case participant ID column to do our predictions. However, about 10% of defendants' information on sentencing length is missing or inaccurate; and 60% of law enforcement unit data is also missing which limits our ability to investigate the factor of courts. Second, we used the Chicago Police Stations dataset that was updated on 10th June 2016 from The City of Chicago's Data Portal. This data includes the specific geographic locations and coordinates of police precincts. The latitude and longitude of the police districts are crucial for creating distribution maps. Still, the data is not complete and contains only 23 police districts from the actual 25. Due to the lack of the name of the neighborhood where each district was located, we added a column to the data processing manually later. Third, we used the JSON file of neighborhood boundaries in Chicago from the same data portal that was updated on 11th July 2018. This file contains geographic information but we believe that there may be a mismatch between the longitude and latitude corresponding to the specific geographical location in the document, which may lead to some errors when drawing the map.

4 Methods

4.1 Data Cleaning and Merging

In order to achieve the predictive power of our machine learning models over the existing data sets, we took several steps to clean the data sets first.

The majority of the independent variables in our models are mainly from the Sentencing_asof0405 data set. We first filtered the data to only where "PRIMARY_CHARGE_FLAG" is true to create a data frame called df1. Then, we selected 2 existing variables from the filtered data set, including To make our model strong enough to perform feasible predictions, we created other 29 variables. During this process, we filtered the values that are NA values or abnormal values to improve the accuracy of our model predictions. For following data cleaning and merging procedures, if not mentioned specifically, new variables were created in the df1 data frame.

4.1.1 Create binary variables

In order to quantify the categorical value of the race variable, we created 4 binary variables that only have True and False values called "is_white_derived", "is_black_derived", "is_hisp_derived", and "is_other_derived". For gender, we created the "is_defendant_male" variable. Since the sentencing data don't have gender information for the judges, we used the gender Detector package to predict the judges' genders to create the "is_judge_male" variable. For court information, we quantified the court name using 1 to 6 to represent every court name of the "SENTENCE_COURT_NAME" variable. Then, we created 6 binary variables, called "is_district[n]_derived", where [n] took on a value between 1 and 6. To study if the change in offense categories would affect sentencing outcomes, we created the "is_changed_offense" variable based on the defendants' original offense categories column and updated offense categories column. Meanwhile, we created the "Incar" variable by setting "COMMITMENT_TYPE" equal to "Illinois Department of Corrections". To quantify the crime classes, we created 10 binary variables of crime types that occur most frequently based on 'UPDATED_OFFENSE_TYPE': 'is_nar_derived', 'is_sex_derived', 'is_wea_derived', 'is_vehi_derived', 'is_theft_derived', 'is_homi_derived', 'is_bur_derived', 'is_bat_derived', 'is_rob_derived', 'is_DUI_derived'. Lastly, to explore the effects of COVID-19 on the judge's judgment, we selected the "SENTENCE_DATE" column and used March 11, 2020, the outbreak day of COVID-19 announced by WHO, as the cut point to create a binary variable called "is_covid".

4.1.2 Create numeric variables

We selected "COMMITMENT_UNIT", "COMMITMENT_TERM" and "AGE_AT_INCIDENT" to recreate a new variable called "senlength_derived" to calculate the defendants' sentence length by year unit. We reset the sentence length of the defendants whose commitment type was natural life as the difference between the age of 100 and the defendant's age at the incident(if their age at the incident was missing, we assumed their age as 20). We recoded the values in "DISPOSITION_CHARGED_CLASS" which contain M, X, 1, 2, 3, and 4 into 6, 5, 4, 3, 2, 1 (from the least severe to the most severe) based on the level of severity of felonies. Moreover, we selected "charge_count" and "age_at_incident" in the sentencing data set to see how would the defendant's age at the time of the incident and the number of charges in each case affect sentencing results. We created the 'Livability' variable by applying the livability score from `https://www.areavibes.com/` to each neighborhood that we manually added to the Police Station data set (The column named "NEIGHBORHOODS". The website calculates a livability index based on the cost of living, nearby amenities crime, employment, housing prices, schools, and user ratings. Finally, we created the 'his_prop' variable by dividing the total number of crime cases in Chicago by the number of cases in each law enforcement unit (neighborhoods).

4.1.3 Merging

First, we selected the "LAW_ENFORCEMENT_UNIT" column in the sentencing data set and filtered out NA values in this column to create a new data frame. We used the count function to count case numbers in each law enforcement unit and added the column of "COUNT" to the data frame. Then, we extracted the district value in the "LAW_ENFORCEMENT_UNIT" to create a new column called "DISTRICT NAME". By doing so, we can merge the data frame with the Police Station data set on "DISTRICT NAME". Then, we used this merged data to inner merge with df1 on "LAW_ENFORCEMENT_UNIT". Last, we created the final data frame for prediction models by selecting "CHARGE_COUNT", "AGE_AT_INCIDENT", "is_changed_offense", "is_black_derived", "is_hisp_derived", "is_other_derived", "is_white_derived", "is_defendant_male", "is_judge_male", "is_covid", "Livability", "his_prop", "Incar", "DISPOSITION_CHARGED_CLASS", "is_nar_derived", "is_sex_derived", "is_wea_derived", "is_vehi_derived", "is_theft_derived", "is_homi_derived", "is_bur_derived", "is_bat_derived", "is_rob_derived", "is_DUI_derived", "senlength_derived", "is_district1_derived", "is_district2_derived", "is_district3_derived", "is_district4_derived", "is_district5_derived", and "is_district6_derived" columns and filtering out NA values in both "CHARGE_COUNT" and "AGE_AT_INCIDENT" columns, values larger than 80 in "AGE_AT_INCIDENT", and values other than "1", "2", "3", "4", "X", "M" in "DISPOSITION_CHARGED_CLASS".

After cleaning the data, the cleaned data set had 47718 rows, and the basic information of important variables we assumed was shown in Table 1.

4.2 Descriptive Analysis and Visualization

We studied the relationship between each variable and each sentencing outcome. To be specific, we did faceted graphs to show each relationship between incarceration/sentence length and individual variables (gender, race, crime class, and courts). We also investigated each relationship between the classification of felonies and the mean incarceration rate/mean sentence length. For visualization, we used Illinois latitude and longitude values (latitude = 41.9811, longitude = -89.5951) and a GeoJSON file of neighborhood boundaries in Chicago to create a Chicago map. Then, we made a heat map of the Chicago map and the latitude and longitude information from our new data frame that we merged with the df1 and the Police Station data set.

Table 1: Description of part of variables in the cleaned data set

Column Name	Min	Max	Mean	Std
AGE_AT_INCIDENT	17	76	33.350664	12.196065
is_black_derived	0	1	0.793307	0.404938
is_defendant_male	0	1	0.889979	0.312920
is_nar_derived	0	1	0.682028	0.465694
is_homi_derived	0	1	0.002808	0.052918
is_rob_derived	0	1	0.033132	0.178983
is_covid	0	1	0.015277	0.122655
Livability	51	76	60.240978	5.471066
DISPOSITION_CHARGED_CLASS	1	6	1.849637	1.222712
Incar	0	1	0.549646	0.497534
senlength_derived	0	82	2.639544	3.018905

4.3 Using Machine Learning Models

Before we put the variables into the prediction models, we converted the values of the binary variables to dummy variables with only 1 and 0 values. Then, we started to identify the feature and the label columns of the models. Since our prediction focused on sentencing results of incarceration and sentence length, "Incar" and "senlength_derived" were our target label columns in the models. We used the train_test_split function to split our original data set into training and test sets (X_train, X_test), with the test data size of 25%. We set the random seed as 42 for the entire process to ensure the accuracy of our prediction models.

For the prediction of incarceration or not, we used four models to make predictions and compare their results, including Logistic Regression, KNN Classifier, Decision Tree Classifier, and Random Forest Classifier. In fact, the logistic regression model is usually used to predict binary variables, and its accuracy to predict nonlinear variables is low. The principle of KNN model prediction is to determine which category x belongs to based on the category of its nearest K points. Our model contains about 30 independent variables and over 30,000 defendants, so it took a long time to run. The decision tree model divides each variable into different branches and predicts the result according to the branch to which the defendant belongs. However, the decision tree model sometimes overfits the results and gets some inaccurate results. Among the four models, the random forest model is more accurate because it creates multiple trees to avoid the problem of overfitting. We concluded this with our results. Further, we improved the model by adjusting the parameters of the decision tree model and the random forest model to achieve a more accurate predictive value. In detail, we set a basic line of max_depth, max_features, min_samples_leaf, min_samples_split, and n_estimators of these models and used the "best_estimator_" function to get the specific value of the best model. We modified the variable that had reached the best utility and change the parameter that has room for improvement, again, based on the value of best_estimator_. After several parameter adjustments, our model finally reached the optimal case.

For the prediction of defendants' sentence length, we chose Linear Models, KNN Regressor, Decision Tree Regressor, and Random Forest Regressor to make predictions. In order to compare the prediction results of different linear models, we chose four different types of prediction, including the linear regression model, Ridge regression model, Lasso model, and ElasticNet model. Although these models can solve the overfitting and irreversibility problems of linear regression, their algorithms are generally not as efficient as those of the decision tree model or the random forest model. As well, we also redefined the parameters of the decision tree and random forest to improve the prediction accuracy. Finally, we compared the results of the eight models.

5 Results

5.1 Relationships between Sentencing Outcomes and Individual Variables

The relationships between individual variables for each category and incarceration/sentence length are presented below. In Figure 1, we observed the defendants whose offense types were battery, burglary, homicide, sex, weapon, and robbery related were more likely to be incarcerated. Male defendants, black defendants, and defendants who were sentenced before COVID-19 were more likely to be incarcerated. As shown in Figure 2, the higher the level of felony the defendants were convicted of, the more likely they would be incarcerated.

From Figure 3, we observed the defendants whose offense types were battery, burglary, homicide, sex, weapon, and robbery related were more likely to be sentenced to a longer sentence length. Among these categories, defendants related to homicide were given particularly long prison sentences. Offenders linked to robbery might also face lengthy jail terms. As shown in Figure 4, the higher the felony the defendant was convicted of, the longer the sentence he would face.

5.2 Geographic Visualization

In order to visualize the geographic information and distribution of different defendants in the Chicago Police Department, we used the "Chicago_neighbors _geojson" file and drew the boundaries of Chicago neighborhoods. Our Choropleth Map and Heat Map, based on the number of cases, show the distribution of the number of cases in the Chicago area. The dark dots of concentrated color in Figure 5 indicated the area with the most cases. This did not entirely explain the high crime rate in the region, but it explained the fact that the courts in this area recorded the most cases. An alternative explanation is that because the original sentencing data set included only a portion of the Chicago Police Department, about 60% of the defendant law enforcement unit was missing, so our heat map might not present accurate information.

5.3 Incarceration Prediction

After training different classifiers to predict whether a defendant will be incarcerated, among the four machine learning models we used, the one with the highest accuracy is the random forest model. However, the logistic regression model has the lowest performance. The accuracy of the refined Random Forest classifier with the changed parameters is 86.3%, a 0.7% improvement from the original 85.6%. Table 4 examined the confusion matrix of different classifiers and showed the

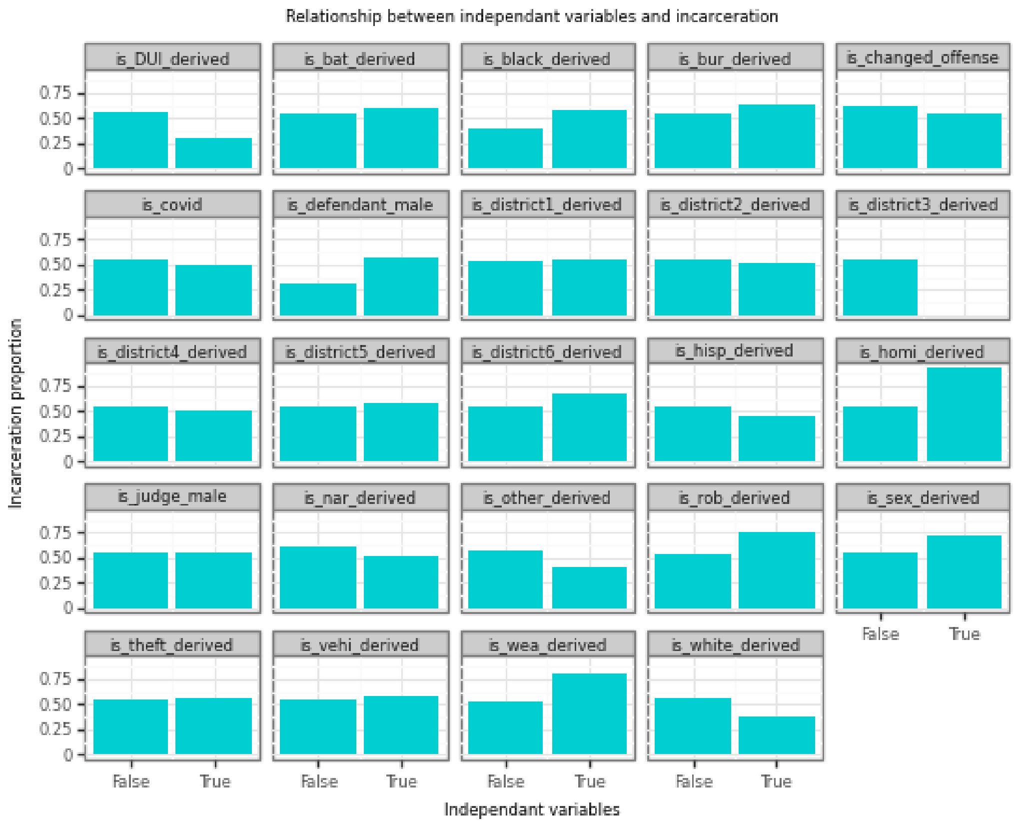

Figure 1: Comparison between the relationships between individual variables and incarceration possibility

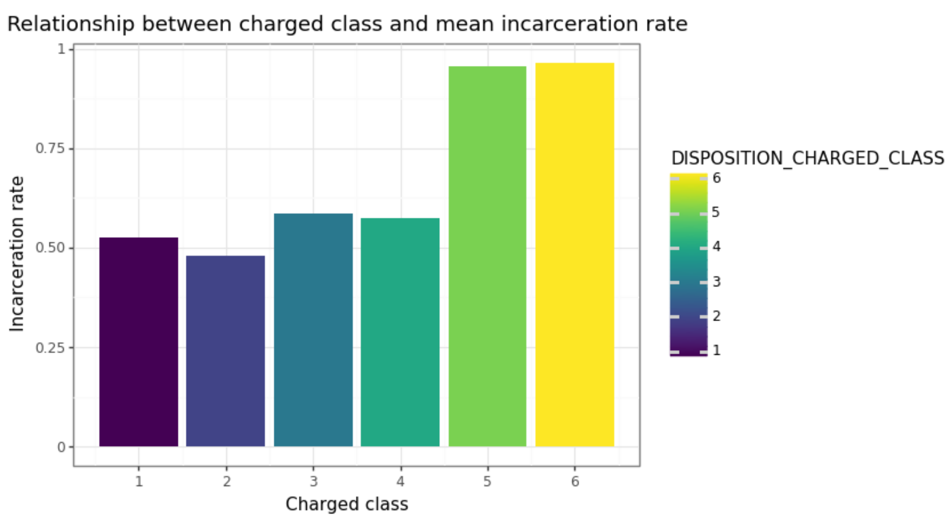

Figure 2: Comparison between disposition charged class and incarceration possibility

Figure 3: Comparison between the relationships between individual variables and sentence length

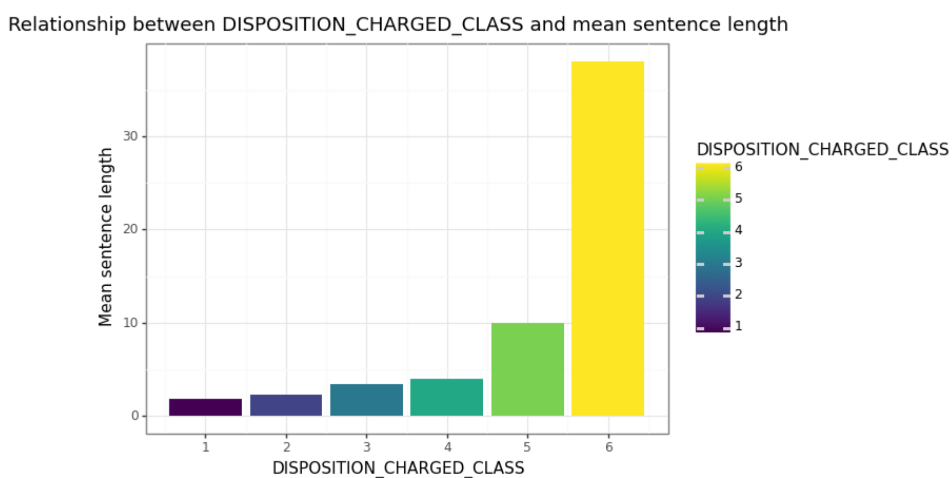

Figure 4: Comparison between disposition charged class and incarceration possibility

results of the best class prediction accuracy of the improved random forest model. The similarity between actual and predicted results proved the relative accuracy of the model. Meanwhile, we created Table 2 that had the accuracy, precision, and recall scores for all models. Since the F-1

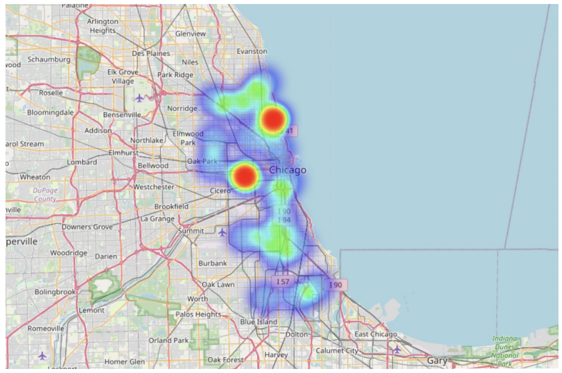

Figure 5: Results: Heat Map of Cases Distribution in Chicago

value balances the precision value and the recall value, we chose the F-1 value other than the mean value as the indicator to measure the prediction accuracy. In addition, we drew the ROC curve of these models, as shown in Figure 6. The larger the area enclosed by ROC curve, the higher the accuracy of the model prediction.

$$F_1 = (\frac{recall^{-1} + precision^{-1}}{2})^{-1} = 2 * \frac{precision * recall}{precision + recall}$$

Table 2: Model Results

Model Type	Accuracy	Precision	Recall
Logistic Regression	0.63	0.63	0.63
KNN Classifier	0.78	0.78	0.78
DecisionTreeClassifier	0.69	0.69	0.70
Random Forest Classifier	0.85	0.86	0.85
Decision Tree Classifier Refined	0.83	0.84	0.83
Random Forest Classifier Refined	0.86	0.86	0.86

We also identified the most important feature that participated in our prediction models. Table 3 shows, in turn, the top 5 most important features involved in the four models. The most important features were the defendants' sentence length, disposition charge class, age, and gender, although each model ranked the features differently. Of these important features, "senlength" played a greater role when determining whether the defendants would be incarcerated or not. Therefore, we believe that the longer the sentence, the more likely the defendant is to be incarcerated. Figure 7 shows an overall comparison of the importance of different model features.

Since these accuracy rates are within the ideal range, the possibility of being incarcerated appears to be predictable when using a binary classifier based on the defendant's information. Compared with the other three models, the random tree performs best with its relatively fast running speed and the highest prediction accuracy.

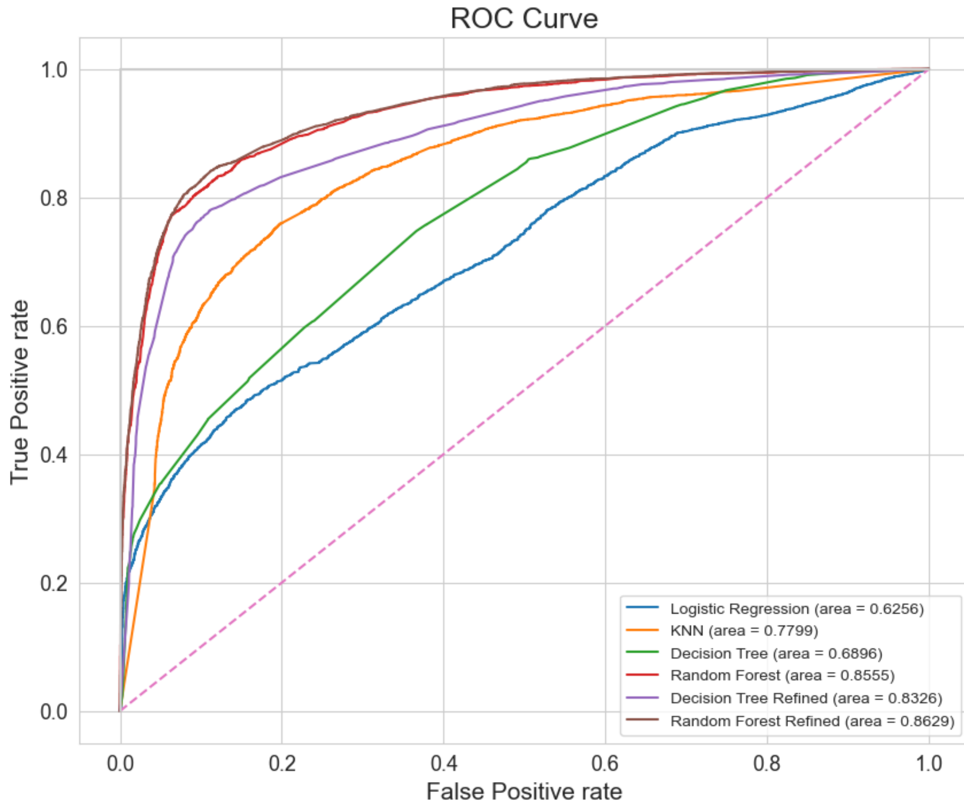

Figure 6: Results: ROC Curve Comparison

Table 3: Most Important Features for Prediction of the Incarceration Condition

Decision Tree	Random Forest	Decision Tree Refined	Forest Forest Refined
senlength_derived: 0.61	senlength_derived: 0.81	senlength_derived: 0.79	senlength_derived: 0.76
is_black_derived: 0.1	CHARGED_CLASS: 0.09	CHARGED_CLASS: 0.06	CHARGED_CLASS: 0.09
AGE_INCIDENT: 0.08	is_black_derived: 0.03	AGE_INCIDENT: 0.04	AGE_INCIDENT: 0.03
is_defend_male: 0.05	AGE_INCIDENT: 0.02	is_black_derived: 0.01	is_black_derived: 0.02
CHARGED_CLASS: 0.05	is_defend_male: 0.01	is_other_derived: 0.01	his_prop: 0.02

Table 4: Refined Random Forest Confusion Matrix

Actual Class	Predicted Class		Class Accuracy
	Not Incarceration	Incarceration	
Not Incarceration	4772	668	0.86
Incarceration	982	5508	0.86

5.4 Sentence Length Prediction

We chose to use regressions to predict sentence length. Unlike what we did in the previous prediction of applying the classifier model, we used the RMSE (Root Mean Squared Error) and R-Squared as indicators to measure the prediction accuracy of our models. RMSE refers to the differences between

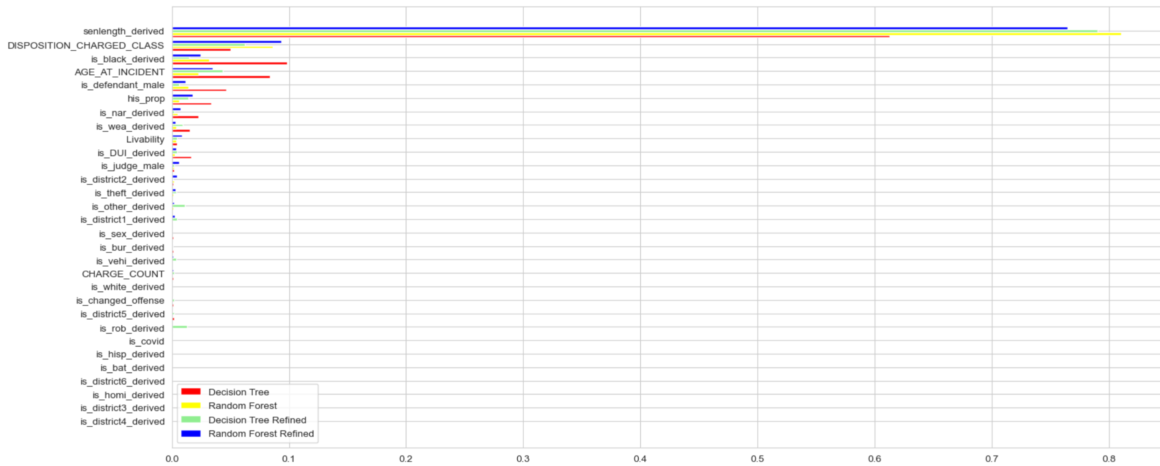

Figure 7: Results: Feature Importance Comparison

values (sample or population values) predicted by a model or an estimator and the observed values, while r squared refers to the variance proportion of the dependent variable in the model explained by the independent variable. As shown in Table 5, both the random forest model and the decision tree model have reached the highest R-Squared value and RMSE.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$

Table 5: Model Results

Model Type	Train RMSE	Train R-Squared	Test RMSE	Test R-Squared
Linear Regression	2.263071	0.430453	2.208297	0.485451
Ridge Regression	2.263106	0.430435	2.208985	0.485131
Lasso Regression	2.263368	0.430303	2.209096	0.485079
ElasticNet Regression	2.263190	0.430393	2.209598	0.484845
KNN Regression	0.860378	0.917678	2.244052	0.468654
Decision Tree Regression	1.825669	0.629338	1.863243	0.633689
Random Forest Regression	1.718697	0.671502	1.837658	0.643679
Refined Decision Tree Regression	1.839367	0.623755	1.870745	0.630733
Refined Random Forest Regression	1.640614	0.700673	1.840671	0.642510

As we did in the incarceration prediction, we also extracted the top 5 important features from our models shown in Table 6. We found that the defendants' charged class, offense type, and age at the incident had a great impact on the defendant's sentence. Among these feature importance, "DISPOSITION_CHARGED_CLASS" played the most important role in determining the defendants' sentence length. Therefore, we believe that the more serious the charges against the defendant, the longer the sentence. In our current data set, homicide, narcotics, and robbery are the three most serious types. Incarcerated people would always receive a longer sentence. At the same time, the age of the defendants at the time of the incident also plays a role in their sentences, possibly because older defendants are not usually first-time felony offenders. In addition, we also drew a comparison graph of the importance of features in the four models, as shown in Figure 8.

Table 6: Most Important Features for Prediction of the Incarceration Condition

Decision Tree	Random Forest	Decision Tree Refined	Forest Forest Refined
CHARGED_CLASS: 0.57	CHARGED_CLASS: 0.52	CHARGED_CLASS: 0.57	CHARGED_CLASS: 0.52
is_homi_derived: 0.22	is_homi_derived: 0.19	is_homi_derived: 0.22	is_homi_derived: 0.16
Incar: 0.09	Incar: 0.08	Incar: 0.09	Incar: 0.07
is_nar_derived: 0.05	is_rob_derived: 0.05	is_nar_derived: 0.05	AGE_INCIDENT: 0.06
AGE_INCIDENT: 0.02	AGE_INCIDENT: 0.04	AGE_INCIDENT: 0.02	is_rob_derived: 0.04

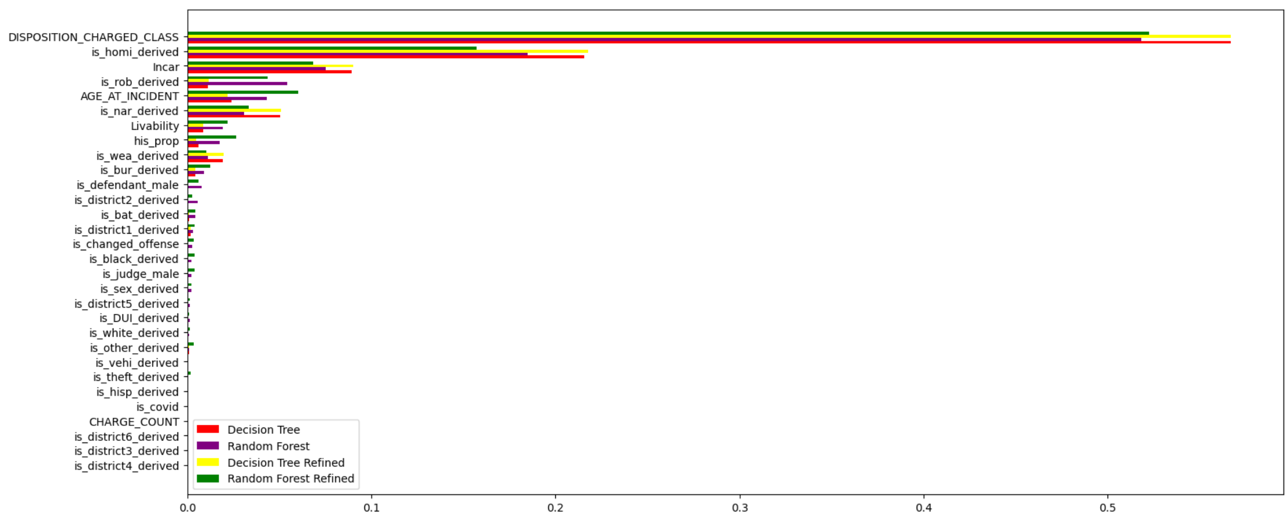

Figure 8: Results: Feature Importance Comparison

6 Discussion and Limitation

Based on the results of our prediction model, we can conclude that our predictions of defendants' incarceration are relatively accurate, with approximately 86% of defendants' incarceration accurately predicted. However, our predictions of the length of sentences fell short of our expectations, with only about 65% of the defendants' sentences being accurately predicted. We believe this comes from the omission of law enforcement unit data in our current data set. There is a huge difference between the number of defendants before COVID-19 and those during it. Thus, the lack of sufficient sample size in the post-pandemic era could affect our predictions. In addition, we need more data to form our prediction sample, such as the defendant's criminal record. Time served and incarceration are affected by complex and fitting factors, and our current run would be improved by adding more potentially relevant variables. We only extracted about 80% of the crime types in the original data and did not take into account the influence of other crime types not included in our forecast range. It is worth considering in future studies whether the same class of crime is charged with the same charge, and therefore how sentence length might vary. Based on specific cases and functions of the models, further research can expand on other machine learning models such as boosting trees and Neutral Network models.

Acknowledgments

Thank you to Professor Rebecca Johnson and the TAs for providing guidance and feedback to our group!

References

- Macmadu, Berk, J., Kaplowitz, E., Mercedes, M., Rich, J. D., & Brinkley-Rubinstein, L. (2020). COVID-19 and mass incarceration: a call for urgent action. The Lancet Public Health, 5(11), e571–e572, from `https://doi.org/10.1016/S24682667(20)30231-0`

- Ostrom, C. W., Ostrom, B. J., & Kleiman, M. (2004). Judges and discrimination: Assessing the theory and practice of criminal sentencing. Final Grant Report to the National Institute of Justice(Grant 98-CE-VX-0008).

- Sawyer, Wagner. (2022). New data: The changes in prisons, jails, probation, and parole in the first year of the pandemic. Prison Policy Initiative. Retrieved December 6, 2022, from `https://www.prisonpolicy.org/blog/2022/01/11/bjs_update`

- Wagner, W. S. & P. (2022). Mass incarceration: The whole pie 2022. Prison Policy Initiative. Retrieved December 6, 2022, from `https://www.prisonpolicy.org/reports/pie2022.html`