Tara Bhardvaj| Hannah Laper |Venkatsh Raj Rudrappa

UMBC  DATA 606

Modeling Bail Outcomes

**Introduction**

On a day-to-day basis, most people do not think about arraignment trials and bail.  It is a topic that does not encourage much thought until there is a reason to think about it.  For example, you may not even know how bail is determined until you or someone you know is a participant in an arraignment.  When a person is arrested for a crime, they have an arraignment in which both the prosecutor and defense provide background information about the arrest and the defendant's criminal history, make statements on their opinion of bail should be, and then the judge determines whether or not to set bail, the bail price, and/or any non-monetary restrictions.  With the judge being the final decider on the bail, it comes to question how fair the setting of bail is.  There are a lot of different determinants that might affect the judge’s decision.  One might immediately think that the crime is the biggest deciding factor of bail, but does gender and race affect the bail set as well?  What about the time of day? How much sleep did the judge get the night before?  Is a judge impartial enough to set a fair and effective bail?  In Malcolm Gladwell’s *Talking to* Strangers (2019), Malcom also questions the effectiveness of a judge determining bail set.  Malcolm references a study by an economist named Sendhil Mullainathan and some computer scientists.  The group took New York City’s data from 2008 to 2013 on the defendant’s age and criminal record and created an artificial intelligence machine to create a list of defendants the machine would release to await trial at home.  The study found that when compared to the actual outcomes decided by judges, the list created by the AI machine contained a population 25% less likely to commit a crime while awaiting trial (Gladwell, 2019).  Further, the AI machine created a risk distribution to categorize the defendants on low, middle, or high risk.  When compared to the actual decisions made by judges, they found that the judges were setting bail for persons throughout the distribution (Gladwell, 2019).  This study shows just how random the bail set can be, and leads to the question who is more effective at setting bail: man, or machine?

This study will explore the New York State Pre Trial data (Division of Technology & Court Research, 2022), an extensive dataset that includes information about the arrest and arraignment of criminal cases in New York.  From this data, the correlation between gender, race and bail set will be analyzed, and then prediction models will be created for predicting whether bail is set, the amount of bail, and non-monetary restrictions set.  These prediction models will be created with the intent of allowing defendants’ and/or their families to be able to enter their information into the models before their arraignment to get an idea of what the bail might be before the arraignment.

**Dataset**

The dataset that is used to perform the analysis was provided by New York State Unified Court System. The dataset contains statewide criminal arraignments, and the data is updated once every six months to add new cases and to update the existing prior extracts. In this dataset, we have various columns that provide details about the cases in which the individual had been charged and the court details which determined whether the individual is charged, or the case is dismissed.

In this dataset, we have details collected in and around New York city which include 63 courts approximately. Arraignment Release decisions may vary from one court to another court and the exact details were also documented in the data. Based on the details available, we have segregated different columns to perform different hypotheses. The most used columns in the dataset for all hypotheses are Gender, Race, Ethnicity, Age at crime & arrest, Arraign Charge Category, Representation Type in the court, Release Decision at Arraign, Bail Amount, and Arrest Type. The total size of the dataset is 194MB - .csv file format which has 284096 rows and 108 columns.

**Data Attributes**

The dataset has a different set of attributes that stand as the reason for which the Arraign Charges cases have been filed and arrested. Out of all available attributes, we have segregated the important attributes closely related to Arraign charge categories in the dataset. Below are the commonly used attributes for identifying the following hypotheses,

* **Gender**: Male, Female, and Unknown; Categorical values
* **Age\_at\_crime**: Age in years; Integer values
* **Race**: Black, White, Unknown, Asian/Pacific Islander, American Indian/Alaskan Native, Other values
* **Ethnicity**: Hispanic, Non-Hispanic, Unknown
* **Offense Date**: Contains date in MM/DD/YYYY format in which the offense has been made
* **Arraign Charge Category**: Categorical values; Has different charges sections for which the case has been filed
* **Court\_ORI**: Alpha-Numeric value; Validates legal access to Criminal Justice Information (CJI) and identifies the agency in all transactions.
* **Judge\_Name**: Contains Categorical values; Name of the judge during arraignment
* **Arrest Type**: DAT & Custody as values; Categorical values
* **Representation Type**: Contains Categorical values; how the arraignment is handled for any case
* **Top\_charge\_at\_arrest**: Contains Numeric and Categorical values; charges made at the time of the arrest
* **Top\_charge\_Severity\_at\_arrest**: Contains Categorical values; Felony, Misdemeanor, Infraction, Violation
* **Bail\_Amount**: Amount paid to get bail; either Cash or credit
* **Bail\_Status**: Yes or No; Denotes the bail status after the arraignment
* **Release Decision at Arraign**: Contains Categorical values; Bail-set, Disposed at Arraign, ROR, Nonmonetary release, Remanded

**Previous Studies**

Previous researchers have attempted to use machine learning to ameliorate the bail process. There is currently very little oversight for judges who seek guidance in determining what bail amounts are appropriate. Consequently, judges must rely on their own judgement of each case to come to an appropriate bail determination. This is a highly complex task as the details of each case are different and it can be difficult for judges to weigh the potential benefits and risks. Studies have also shown that environmental factors play a role in judicial outcomes, with one study finding that at the beginning of the day judges opted for parole in 65% of cases but that by the end of the day they opted for parole in 0% of the cases. After the judges took a break, their parole approval decisions went back to the initial 65% in favor of parole (Kleiner,2011). This was attributed to decision fatigue, a phenomenon wherein as people get tired of making decisions, they begin to default to simple answers that may not account for the full scope of information presented (Kleiner,2011).

In bail hearings this may result in wildly differing bail decisions being handed down due to variance in the judge’s mental state as opposed to differences in the case. In order to address this issue, New Jersey state adopted an algorithmic risk assessment model that returned a level of dangerousness for each defendant (Simonite, 2020). The PSA risk assessment model has been hailed by some as a way to take the bias out of bail decisions (Simonite, 2020). While critics, note that the data used to build the algorithm may reflect the existing racial and socioeconomic inequities in policing and crime (Simonite, 2020). The move to adopt these risk assessments has been made in order to minimize the use of cash bail wherein a defendant is released if they are able to pay the bail (Simonite, 2020). Opponents of cash bail state that it unfairly punishes those of a lower socio-economic status as their inability to pay the lump sum may result in incarceration (Simonite, 2020). In practice, in cases where PSA tools have been implemented some areas have found that the tools exacerbated racial disparities in sentencing, with White defendants receiving bail more often than Black defendants (Simonite, 2020). However, in some cases the results were positive.

Kentucky implemented the PSA tool and found that when judges considered the algorithm derived measure of dangerousness that more people were given bail and that more minority defendants were given bail as compared to similar cases in the past (Covert, 2020). In this time, there was a 13% increase in defendants released on their own recognizance, meaning that they did not have to post bail or abide by any non-monetary conditions (Covert, 2020). However, once judges were no longer obligated to consider the PSA findings disparities in bail returned (Covert, 2020). Specifically, judges were most likely to override moderate risk scores for Black defendants and grant them cash bails (Covert, 2020). The benefits of the assessment were not all entirely positive. The scores gave defendants without a permanent address a higher risk rating regardless of the seriousness of their crime (Covert, 2020). The basis for this decision is the prevailing idea that they will fail to appear for their court date (Covert, 2020). However, this is not necessarily a foregone conclusion, as many people who do not have a permanent address are able to diligently attend court with the support of organizations like the Bail Project (Covert, 2020). In addition, the risk assessments themselves may not be all that meaningful. In Cook County Illinois, 99% of high-risk defendants who were granted bail were not charged with a violent crime between their release date and their trial date (Covert, 2020).

Determining bail is a delicate mix of weighing public safety, judicial time, and the wellbeing of the individual and their community. The calculations are complex and sometimes harrowing as judges must determine if the benefits to some outweigh the dangers to all. Using ML to adjudicate these decisions does not always lead to a fairer version of justice. It can easily be implemented in a way that replicates inequities in our society and reinforces biases in our system. In our research, we aim to simply model bail decisions to provide a model that will, for better or worse, replicate the judicial decisions that are currently being made.

**Hypotheses/ Goal of Project:**

This project will explore two different aspects of the dataset.  The first goal is to analyze the dataset to explore whether or not there is a correlation between race and bail set, or gender and bail set.  We predicted that there is at least a 10% difference between the frequency of bail set, price, and nonmonetary restrictions within each sub-population of the dataset of all people in the dataset.  The second goal of the project is to create three prediction models that predict whether bail is set, the bail set amount, and whether there will be non-monetary restrictions.  The intent behind the prediction models is to make an interface where people can enter their information: age, gender, race, charge, charge severity, prior criminal acts, etc. and their bail will be predicted.

**Data Exploratory Analysis**

We conducted exploratory data analysis to develop insight into the dataset.

First, we created a general overview of the data. This included cases over time, the age range, and the racial distribution of arrests as well as a distribution of the types of crime.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure 1**

Then we dove deeper into demographic details analyzing the percentage of cases where bail was set compared to the percentage of cases where bail was not determined (fig.2). As well as the percentage of cases where bail was set (fig. 3). We also looked at the breakdown of cases by race and gender. We found that Black and White women are charged at similar rates, but more Black men were charged than White men (Fig. 4).

Chart, pie chart

Description automatically generated

**Figure 2**

Chart

Description automatically generated with low confidence

**Figure 3**

Chart, treemap chart

Description automatically generated

**Figure 4**

Next, was an examination of race and charge. Analysis showed that for the two most common charge categories; criminal contempt and assault, the distribution of felonies to misdemeanors was roughly half. In addition, Black defendants were charged nearly twice as much as White or Hispanic defendants were. Furthermore, Black defendants were 11% more likely to be charged with a felony than White defendants. The distribution of felonies to misdemeanors was 2:1 for Black and Hispanic defendants and around 1:1 for white defendants (Fig. 5). Chart, timeline

Description automatically generated

**Figure 5**

Next, we conducted an analysis of differences by geographic region. We found that certain districts were overrepresented with regards to crime statistics and arrests. In addition these counties have different demographics as evidenced by the various racial mixes represented and the varying average bail amounts(fig 6). Graphical user interface, chart

Description automatically generated

**Figure 6**

Lastly, we examined the distribution of bail amounts and the varying medians and averages for each crime type. This showed that the average bail value is typically higher than the median value. This represents 2 facets of our data. First, that the most common bail amount is $1 which is extremely low, and second, that for many of these charge categories there were a handful of cases with exceedingly steep bail amounts. This creates a situation where the data is skewed towards larger values and the median value is artificially lowered by the plethora of low bonds( figs 7&8).

Bar chart

Description automatically generated with low confidence

**Figure 7**

Chart, bar chart, histogram

Description automatically generated

**Figure 8**

**Methodology**

**Predict Bail set or not:**

According to the dataset documentation, we have so many attributes which were the reasons and factors for the arraign charge imposed on any individual in and around New York City. To predict Bail status, we have segregated a few attributes from the main dataset that can contribute more likely to release decision at arraign. The attributes selected are 'Gender', 'Age\_at\_Crime', 'Race','Ethnicity','Offense\_Date', 'Arraign Charge Category', 'Representation\_Type','Release Decision at Arraign', 'County\_Name'. The attribute in the dataset contains so many missing values and it is replaced with appropriate values using math functions, respectively.

Mostly the attributes selected has categorical values which represent various traits. Regardless of what the values is used for, the major challenge is determining how to use the data for any analysis. Though we have many Machine learning models which can support and provide better results with Categorical values without any manipulation there are many more algorithms that do not. Thus, we have used a different approach for different attributes to convert categorical values to numerical values. One of the most used methods for converting Categorical to Numerical is the LeaveOneOut Encoders from the category\_encoders package in python.

Timeline

Description automatically generated

**Figure 1**

Also, to predict Bail status with respect to Age as one of the factors, we have grouped the age into different groups and labeled them as 0,1,2,3. Similarly, the offence year is also grouped into four decades as 0,1,2,3 so that we can clearly view that this factor had any contribution to bail status during the arraignment. One of the essential attributes in the dataset is Arraign Charge Category. Here we found top crimes that were charged for arrest and converted that into Numerical values using LeaveOneOut Encoders. To get the exact hypotheses result, we have used the attribute Release Decision at Arraign in which the values are converted into numeric as Bail\_set to 1 and rest other reasons for remanded as 0 respectively.

Once the above steps are performed, we have a dataset (final\_baildf) with only numerical values and then normalized the same to measure all the attributes on the same scale. Normalized data will be easy to construct and run various Machine Learning Models.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure 2**

**Predicting Bond**

Bail amounts are set at the discretion of the judge. This means that they can range anywhere from $1 to hundreds of millions of dollars. In order to prepare the dataset, the years selected were narrowed to 2018-2021. In addition, the bail amount was constrained to ±2 standard deviations from the mean bail amount(fig1). Once this had been done the number of records was constrained to 36,540. Furthermore, the range of bail amounts went from $1 to $999,999,999. Initially, a linear regression model was applied to predict bail amount. The model used the charge category, race, gender, and violence level of the crime to predict bail amount. This model did not produce good results with mean squared error values in excess of 4000. Indicating that the model was not well fitted. Therefore, the bail amount was converted to bail percentiles using the pandas rank command(fig2).

Timeline

Description automatically generated

**Figure 1**

Chart, scatter chart

Description automatically generated

**Figure 2:** Bail amount vs. Bail Percentile

Here we can see that the majority of bail amounts are below $100,000. However, there are outliers that are still within two standard deviations but still an order of magnitude larger than the median value.

The model was then fitted on the data and the result had a much lower mean squared error. After this initial modeling, the features were refined so as to provide more information. The selected features were Race/Ethnicity, Court ORI, Gender, and Charge. The charges, race, and court ORI were one hot encoded as they are categorical variables. A linear regression model was then constructed. Linear regression models are linear models that assume that the data can be modeled using a line. Our data was bimodal and therefore the linear model did not perform well. The model had an R2 value of -1.325 x1022 and a mean squared error of 1.103 x1021 **.** The Decision Tree Regressor performed better with aR2 value of 0.45 and a mean squared error of 0.44. This was surprising as this model had fewer features. The K Neighbors Regression in this case was similar to the results of the same model run on the feature rich dataset with an R2 of 0.452 and a mean squared error of 0.045. Lastly, a Support Vector Regression was run with an R2 of 0.0.51 and a mean squared error of 0.04.

In order to improve performance, additional features were added to improve explainability and performance. The features used were prior felonies, prior violent felonies, prior misdemeanors, age, court ORI (court identification number), gender, race, and charge which includes the law and section of the law that was violated. The age, and priors’ variables were preprocessed using standard scaler. The Linear Regression model still performed poorly, with an R2 value of -2.2 x1014 and a mean squared error of 1.8 x1013. Next, a Decision tree regressor was run. With this model, the R2 value was 0.17 meaning that around 17% of the model’s results could be attributed to the input variables. The mean squared error was 0.067 which is low, but the model still did not produce good results. A K Neighbors Regression was run. That model performed much better with an R2 value of 0.44 and a mean squared error of 0.045. The Support Vector regression returned an R2 of 0.53 and a mean squared error of 0.03. This was the best performance of any model. Furthermore, the R2 value suggests that only around 53% of the model can be explained through the input variable. This suggests that there are features that could be added to the dataset to get a more robust model performance.

The bail amounts were then ranked from lowest to highest and the data set was divided into equally sized quantiles. A logistic regression was run on the multiclass labels of very low, low, medium, and high. This had the following results.

* Recall Score: 0.5045612114577632
* F1 Score: 0.49323880937036235
* Precision Score: 0.49304877929031826
* Accuracy Score: 0.5045612114577632

This suggests that even in the case of a classification problem, the model can only correctly predict the class label around 50% of the time.

**Predicting if bail will be set**

Before we build the Machine Learning Model on the tidy dataset obtained, we decided to find the correlation of different features/factors with Bail status.

Graphical user interface, Teams

Description automatically generated

**Figure 1**

From the heatmap, we can see that Arraign Charge Category is highly correlated to Bail status, followed by County in which the arraignment has been charged. Now, we will build models with the above-obtained data to confirm if the dataset values also confirm that bail decisions are majorly decided with respect to crimes that were committed and county.

To finish up the data pre-processing to provide input to the models, we will split the dataset into two datasets training and testing data. Here, we have divided the dataset as 30% and 70% for Testing and Training, respectively.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure 2**

**Modelling:**

Here, we have performed three models with a different algorithms for the first hypotheses. We will compare the results of all three models and tune the best fitting model to provide more accurate results.

**Baseline Model:**

We started the training of the dataset with Dummy Classifier as it uses simple rules to create predictions. We build this model as a baseline to compare with other models' results.

Text

Description automatically generated

**Figure 3**

Here, we have obtained 71.18% accuracy which is the baseline, and no other model can provide accuracy less than this.

**2. Logistic Regression Model:**

The supervised learning classification algorithm logistic regression is used to predict the likelihood of a target variable. When the target variable is categorical, mostly efficiently used algorithm is Logistic regression.

Accuracy Score: 72.4%

Precision Score: 73.9%

F1-Score: 83.03%

The accuracy of Logistic regression is 72.4% which is above the baseline model. Also, we have a good precision value which confirms the proportion of data that was predicted positive is positive.

**3.Gaussian Naïve Bayes Model:**

Naive Bayes is a simple method for building classifiers, which are models that give class labels to problem cases represented as vectors of feature values, with the class labels selected from a finite set. Given, our labeled training and test data the model produced the below results.

Accuracy Score: 71.1%

Precision Score: 74.1%

F1-Score: 81.7%

The accuracy score got from the Gaussian Naïve Bayes is less than the Logistic Regression Model. Also, the F1 score is less than the Logistic regression model.

**4. K-Nearest Neighbors:**

K-Nearest Neighbors (KNN) classifies fresh data by locating the k-number of closest neighbors from the training data and deciding the class based on most of its neighbors. The model is provided with a k-value of 7 to test and train our data and it produced the below values.

Accuracy Score:76.4%

Precision Score: 79.7%

F1-Score:84.4%

The accuracy obtained from the KNN algorithm is the best so far and it is greater than the above models. The Precision score had increased a lot comparing other models. So, we will tune the parameters to get better results.

**Model Comparison:**

As we have built different models, we will compare the results with graphs to understand the best fitting model with our dataset. We will have two charts one that compares the scores predicted with the models and the other to compare ROC curves from the model.

Chart, bar chart

Description automatically generated

**Figure 4**

From the above chart, we can say that the K-NN algorithm is the best fitting model for our dataset except for the Recall value. So, we will choose the K-NN model further to tune the parameter to obtain better scores and predictions.

In the K-NN algorithm, we have tuned the n\_neighbors value to 5 and found that the model provides better results compared to previous results.

Text

Description automatically generated

**Figure 5**

Further, have modified the split of the dataset into testing and training as 20% and 80% respectively. Fed the model with the new test and train dataset and the model provided a great accuracy score, F-1 score, and precision score above 80%.

Text

Description automatically generated

**Figure 6**

**Predicting Amount**

**Predicting NMR**

Non-monetary restrictions (NMR) are restrictions assigned by the judge at the arraignment.  This can include things like house arrest, mandatory participation in a program (ie. Alcohol or Narcotics Anonymous), maintaining employment, surrendering a passport, etc.  In the dataset, information can be gathered on whether or not a person is assigned a NMR, and what that NMR is.  To better understand the NMR data, brief analysis is created before creating prediction models for NMR.  About 10% of the dataset population receives an NMR as seen in Fig 1.  Fig 2 and Fig 3 show the percentage of persons with NMR by gender and race respectively

Chart, pie chart

Description automatically generated

**Fig 1**

|  |  |
| --- | --- |
| Chart, bar chart  Description automatically generated  **Fig 2** | Chart, bar chart  Description automatically generated  **Fig 3** |

As seen in Fig 2, 7% of females have an NMR while 10% of males have an NMR.  Of the people whose race is known, persons identifying as Black have the highest NMR percentage at 12%, while American Indian/Alaskan Native has the lowest NMR at 6% as seen in Fig. 3.  For both gender and race, the percentage of each sub-population that receives an NMR are within 10% of each other.

Predicting NMR at a person’s arraignment is a binary classification task.  To prepare the data for classification, categorical data that was less than 10 categories was prepared using one hot encoding.  Categorical data with more that 10 categories was transformed using leave one out encoding.  The features of the data were also scaled using sklearn’s standard scaler.  A logistic regression, naïve bayes, k-nearest neighbor, SVM, Random Forest model , and neural network were created to classify NMR at arraignment. Parameters of the models were explored and optimized using Grid Search.  Accuracy, precision, and recall were observed with the goal to determine which model has the highest accuracy.  Model runtime was also observed with each model created.  Runtime will vary between devices being used, but was worth noting as all models were run on the same device.  Results are shown in the figure below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Logistic Regression | Naives Bayes | KNN | Random Forest | Neural Network |
| Accuracy | .71 | .83 | .90 | .64 | .90 |
| Precision | .21 | .22 | .37 | .19 | - |
| Recall | .74 | .28 | .14 | .86 | - |
| Run time | 4 min 37 sec | 2 min 27 sec | 11 min 47 sec | 6.75 sec | 56 min 27 sec |

The K nearest neighbors model and the Neural Network has the highest accuracy at 90%.  The SVM model was not included in the results because its runtime exceeded an hour, and timed out on my device.

Predicting NMR at a person’s arraignment is a binary classification task.  To prepare the data for classification, categorical data that was less than 10 categories was prepared using one hot encoding.  Categorical data with more that 10 categories was transformed using leave one out encoding.  The features of the data were also scaled using sklearn’s standard scaler.  A logistic regression, naïve bayes, k-nearest neighbor, SVM, Random Forest model , and neural network were created to classify NMR at arraignment. Parameters of the models were explored and optimized using Grid Search.  Accuracy, precision, and recall were observed with the goal to determine which model has the highest accuracy.  Model runtime was also observed with each model created.  Runtime will vary between device being used, but was worth noting as all models were run on the same device.  Results are shown in the figure below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Logistic Regression | Naives Bayes | KNN | Random Forest | Neural Network |
| Accuracy | .71 | .83 | .90 | .64 | .90 |
| Precision | .21 | .22 | .37 | .19 | - |
| Recall | .74 | .28 | .14 | .86 | - |
| Run time | 4 min 37 sec | 2 min 27 sec | 11 min 47 sec | 6.75 sec | 56 min 27 sec |

The K nearest neighbors model and the Neural Network has the highest accuracy at 90%.  The SVM model was not included in the results because its runtime exceeded an hour, and timed out on my device.

Once NMR is determined as assigned, a prediction model must be created to determine the specific NMR(s).  To begin, the model will be based on only data where an NMR exists, so all rows with no NMR will be removed from the data frame.  The dataset contains 15 different types of NMR, however 8 of the NMR have less than 1% occurrence in the dataset.  With less than 1% frequency, there is not a sufficient amount of data to predict these NMR restrictions.  Therefore, the prediction model will only classify whether or not you have 1 of the 6 NMR with a greater than 1% frequency in the data.  Below is a breakdown of the frequency of each NMR:

|  |  |
| --- | --- |
| **Non monetary Restriction** | **Frequency of occurrence in dataset** |
| Pretrial Supervision | .44 |
| Contact Pretrial Service Agency | .52 |
| Electronic Monitoring | .05 |
| Travel Restrictions | .01 |
| Passport Surrender | .00 |
| No Firearms or Weapons | .06 |
| Maintain Employment | .00 |
| Maintain Employment | .00 |
| Maintain Housing | .00 |
| Maintain School | .00 |
| Placement in Mandatory Program | .00 |
| Removal to Hospital | .00 |
| Obey Order of Protection | .01 |
| Family Offense | .00 |
| Other NMR | .16 |

Predicting the type(s) of NMR is a multi-output classification model.  Meaning that for each input, we want to predict 6 different binary outputs.  We created 6 different multi-output models to compare for the highest accuracy.  We created a K nearest neighbor, mlp classifier, radius neighbors classifier, random forest classifier, and a neural network.  The results are as follows:

|  |  |
| --- | --- |
| Model | Accuracy |
| Decision Trees | .48 |
| K Nearest Neighbors | .40 |
| MLP Classifier | .45 |
| Radius Neighbors Classifier | .40 |
| Random Forest Classifier | .44 |
| Neural Network | .66 |

As seen in the table, the Neural Network model had the highest accuracy of 66%.  More data would be beneficial in improving all of these models, since NMR restrictions only account for about 9.5% of the original dataframe.  Further, with more data we would be able to predict the 8 NMR options removed due to lack of data. 

**Conclusion**

From this dataset, we were able to create successful models for each of our three goals: predicting bail, bail amount, and non-monetary restrictions. The predicting bail model had an 80% accuracy, the predicting bail amount had 53% accuracy, the model predicting whether there was NMR had 90% accuracy, and predicting the specific restrictions had an accuracy of 66%. To improve these models, we would like to collect more data, continue to fine tune parameters, and update the data used as more cases come in. These models could be helped by expanding the amount of data so as to ensure that the datasets are balanced. In addition, more robust feature sets that included more information regarding each specific case and defendant would be beneficial as more features would likely account for some of the variance in outcome that our models could not capture.

To expand the project, we would like to create prediction models for other state judiciaries and compare the bail data available to find trends between states. The purpose of these models is to use them to create an interface that people can interact with to predict bail based on given conditions: age, gender, race, arrest type, prior criminal arrests, etc. The intent of making this interface is to help people understand what to expect as their bail at an arraignment, and to better inform the population about bail setting.

**References**

1. Covert, B. (2020, July 12). *How a bail reform tool failed to curb mass incarceration*. The Intercept. Retrieved May 16, 2022, from https://theintercept.com/2020/07/12/risk-assessment-tools-bail-reform/
2. Division of Technology & Court Research. (2022).  Pretrail Release Data [CSV]. Retrieved from<https://ww2.nycourts.gov/pretrial-release-data-33136>
3. Gladwell, M. (2019). Getting to Know der Fuhrer. In *Talking to Strangers* (pp. 36–43). essay, Little, Brown, and Company.
4. Kleiner, K. (2011, September 1). *Lunchtime leniency: Judges' rulings are harsher when they are hungrier*. Scientific American. Retrieved May 16, 2022, from https://www.scientificamerican.com/article/lunchtime-leniency/
5. Simonite, T. (2020, February 19). *Algorithms were supposed to fix the bail system. they haven't*. Wired. Retrieved May 16, 2022, from https://www.wired.com/story/algorithms-supposed-fix-bail-system-they-havent/