

## Checkpoint 4 - Findings

Findings pertaining to the time series and moving average model can be found integrated into the “quick\_timeseries.ipynb” notebook. Challenges presented via pursuing a time series model and a moving average model led our team to pursue other opportunities to implement machine learning models.

We begin our new implementation with a new dataset comprising the year, beat ID, allegation ID, gender of the officer and complainant, race of the officer and complainant, and the final finding for the particular allegation. In particular, cases were limited to allegations that resulted in “Not Sustained” (NS) or “Sustained” (SU). Furthermore, cases that had missing information were removed, as well as duplicates. These actions resulted in a dataset of 9,173 rows.

Since most of these are nominal (categorical) variables, and not ordinal, a series of indication variables were used to indicate the presence or absence of a value. For example, `beat_id_6` equals 1 if the beat ID for a given allegation is 6. Conversely, `beat_id_6` equals 0 if the beat ID is not 6 for a given allegation. These substitutions were made because unlike ordinal variables, nominal variables don't have any relation from one value to another. For example, `beat_id = 6` is not inherently better or worse than `beat_id = 5`. After dropping the column for allegation IDs, the data frame had 292 columns. In order to balance the dataset such that there were equivalent cases that resulted in sustained findings to that of not sustained, we down sampled the cases that ended in “not sustained.” These actions led to a data frame of 5,622 rows and 292 columns.

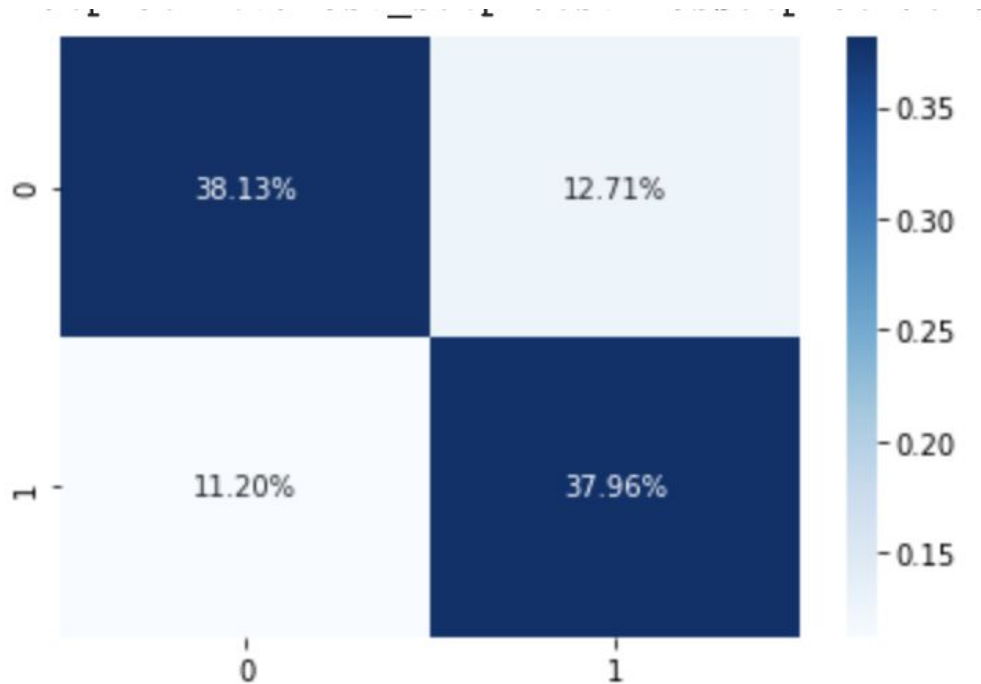
From this dataset, we implement a logistic model. The log model relies on the sigmoid function to output a probability and classify inputted data as one of two outputs (in the case of binary classification). An alternative form of the logistic regression function is  $\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$ . Using this formula, one can further analyze variables of significance similarly to how one might do so with linear regression.

In conducting dimensionality reduction, our group ran into issues with variable significance. Specifically, Every variable was computed as being statistically insignificant. This is counterintuitive and impractical to the aim of our analysis. A model where all the coefficients are insignificant is a model that consists only of the intercept. As such, we proceeded with no dimensionality reduction. That said, we can analyze the relationship of each variable to the outcome SU/NS. For example, `Complainant_Race_White` has a strong positive relationship with a complaint being sustained. Alternatively, `Complainant_Race_Black` has a negative relationship with a complaint being unsustained. There are certainly potential confounding variables that have not been included in the model, but we did seek to account for basic ones such as complainant gender, beat of alleged origin, and officer biographical information. Further analysis is needed to determine if this can lead to conclusive evidence as to the role race, gender and location play in a sustained vs not sustained allegation. For example, the allegation type is another potential confounding variable.

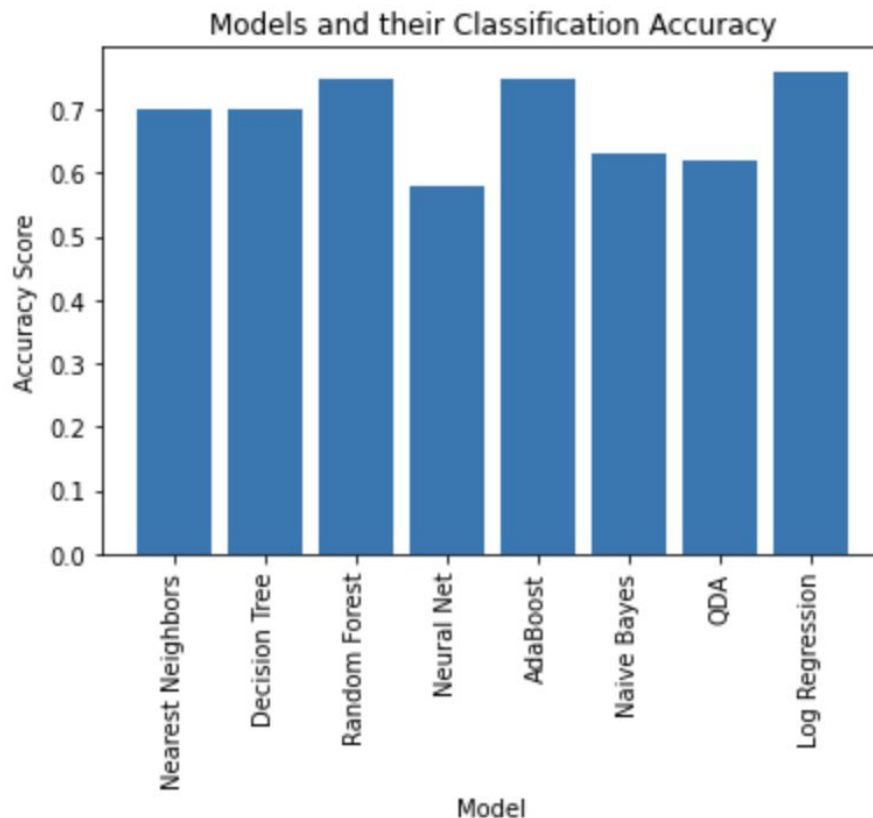
Other findings included a positive relationship between certain beats such as beat 2 and 3 and an allegation being sustained, as well as a negative relationship between beat 1 and final allegation result. Asian/Pacific Islanders had a positive relationship with sustained finding while Native American/Alaskan Native and Hispanic had negative relationships. Further correlations can be found via indexing the regression results.

## Checkpoint 4 - Findings

The result of our model can be visualized with a confusion matrix and an accuracy report. In the confusion matrix, our model performed relatively equal with classifying true positives and true negatives. The algorithm correctly predicted approximately 38% of the data as not sustained and 38% of the data as sustained. Consequently, approximately 24% of the data is misclassified. Both sustained and not sustained classes have decent precision, recall and f1-scores (mid 70s).



Due to these mediocre results, our group attempted to implement other models to test for specific model performance on the data. The accuracy scores for each applied model can be viewed below.



The result of this study demonstrated that logistic regression was the best choice (by accuracy). Similarly, since logistic regression is one of the more simple models, it is therefore the best overall model for the given circumstances.

Our logistic model can inevitably be used to predict the likelihood of a case being sustained or not sustained based on certain features of the case. Consequently, variable analysis can indicate bias and room for growth in the response to certain allegations. More specifically, allegations made by black complainants might experience negative bias while allegations made by white complainants might experience positive bias. As such, black complainants ought to receive further attention equivalent to other races. This is one example of a conclusion we can draw from the log model.