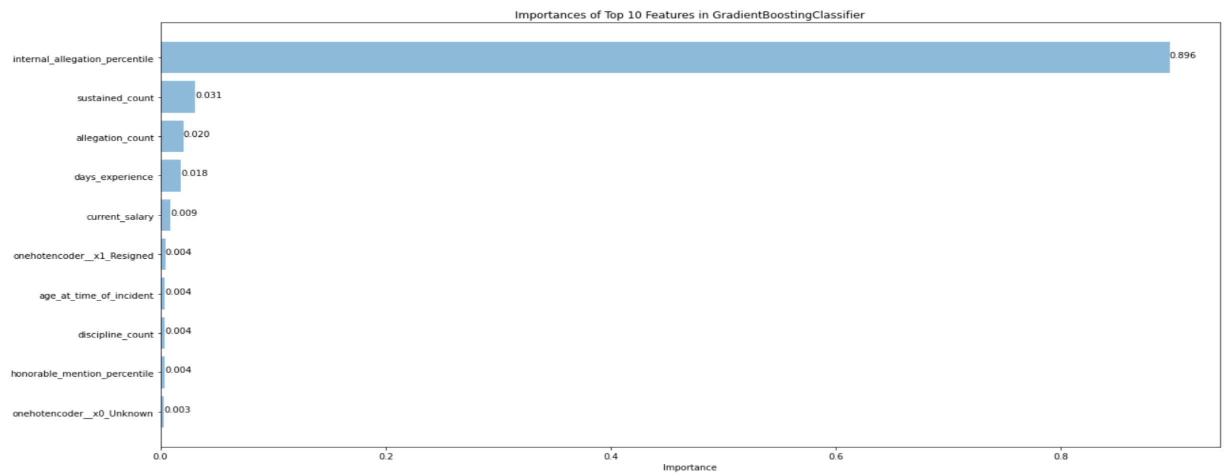
Checkpoint 4

Task #1 (Supervised): Could officer complaints be a 'canary in a coal mine' that could warn about future behavior? In other words, given an officer-filed allegation and the demographic, rank, and other information about the alleged officer, can we predict whether that officer will receive a civilian complaint within 2 years? Specifically, we will collect information about the officer-filed allegation (categories, days, month, victim's gender, age, race, etc.) and information about the alleged officer (rank, age, years in the force, race, etc.). Then, we will look at the next 2 years of these officers to see if they receive a civilian complaint. Using these columns, we can train a supervised model to predict whether the officer will get a civilian complaint in 2 years.

In this supervised ML task, we aim to use officer-filed allegation and other information from the alleged officer to predict whether they would receive a civilian allegation in 2 years. The features we decided to use are basic information about the allegation, such as category, and information about the officer such as their salary, age at the time, and their allegation percentiles. We then train several different models to look for the best one using different models and cross-validation grid search. These models include logistic regression, decision tree, random forest classifier, gradient boosting classifier and neural network. The best model is the gradient boosting model with the following parameter: {'max_depth': 1.0, 'min_samples_leaf': 0.01, 'min_samples_split': 0.01, 'n_estimators': 200}. Our model is able to reach approximately 81% accuracy, with an F1 score of 0.44. The model predicted the negative class (no civilian complaint in 2 years) with 82.8% accuracy, and the positive class with 68.1% accuracy.

We also looked at the importance of various features in the model. Most features had an importance of zero and were not useful to the classifier, and we show the top 10 most important features in the chart below. It's possible that there are confounding factors that could explain the high importance of features like internal allegation percentile, but the presence of a variety of features on this list shows that the prediction is not entirely dependent on just a few features. (Note that the large size difference in importances is just an artifact of the calculation method, based on height in the decision tree, so relative importance is more important than absolute).



These results suggest that one can predict with decent accuracy whether an alleged officer will receive a civilian allegation in 2 years, based on information about the officer and the details of the allegation. Our hypothesis was that a given officer-filed allegation could be used as a "canary in a coal mine" to warn that an officer may receive civilian misconduct allegations in the future, and the moderate success of our model is evidence that supports this hypothesis. Even though the accuracy on the positive cases is only 66%, this still means that an officer allegation can point to future civilian allegations with some accuracy. For future work, since we train our model on both the officer's information and the allegation details, we could separate the two to better understand whether the officer allegation is more predictive than the officer themselves.

Task #2 (Unsupervised): We want to use machine learning to cluster officers and find insights regarding the types of officers who have complaints filed against them by another officer. We will use all the same information as in previous sections which are specific traits that identify an officer. We will also use unsupervised learning to understand which traits are the most significant when determining what leads to a complaint.

In this ML we used K-means clustering in an attempt to group officers who have complaints filed against them by other officers. The features that we decided to cluster on were the same as our previous interactive data visualization. This included: age, gender, years of service, race, number of civilian allegations, and number of awards acquired. Traditionally, K-means clustering is performed on numerical data but features such as gender and race are categorical. To tackle this, we use a library called K-Modes which allows for the clustering of categorical variables. It works by defining clusters based on the number of matching categories between data points (https://github.com/nicodv/kmodes).

We tested different values of k for the clustering and using the elbow method we found that k=3 is the optimal number of clusters. We then used a chi-squared test with the predicted labels to see which features were the most influential. We found that age, the number of civilian allegations a police officer has accrued, and their years of service are the most telling when it comes to separating the clusters.

The number of officer-filed allegations is actually not very significant when it comes to clustering, which is surprising considering that the number of civilian allegations is important. This may be because civilian allegations are more common and thus are a greater tell of whether an officer will receive an officer-filed complaint behind the scenes. The converse may also be true where if an officer has been reported for misconduct among their peers, they are more likely to act similarly to the public.

These results support our visualizations from the last checkpoint where initially we could see that white police officers and males accrued the most officer-filed allegations. After normalizing the data, however, we found that it was actually Native American officers who get the highest average officer-filed complaints. We can now be more confident that race and gender are not huge factors when trying to determine whether a police officer will obtain an officer-filed complaint. We can focus more on the quantitative variables such as age and allegation count.