ANALYSIS OF ARRESTS AFTER TERRY STOPS IN SEATTLE

NAME: BENNIE KIBET

BUSINESS UNDERSTANDING

INTRODUCTION

In Terry v. Ohio, a landmark Supreme Court case in 1967-8, the court found that a police officer was not in violation of the "unreasonable search and seizure" clause of the Fourth Amendment, even though he stopped and frisked a couple of suspects only because their behavior was suspicious. Thus was born the notion of "reasonable suspicion", according to which an agent of the police may e.g. temporarily detain a person, even in the absence of clearer evidence that would be required for full-blown arrests etc. Terry Stops are stops made of suspicious drivers.

Therefore, to help the public department make reasonable judgements when prosecuting such cases they would like to have a model that identifies such cases to plan on how to deal with them.

PROBLEM STATEMENT AND OBJECTIVES

Problem Statement

- To identify whether arrests made after a Terry Stop, has undergone the procedure of being stopped on suspicious behaviour and frisked, or if other factors play a role in the arrest.

Objectives

1. To predict whether an arrest was made after a Terry Stop.

2. Arrest given information about the presence of weapons.

3. The time of day of the call.

FEATURES IN DATASET

Features in the dataset include;

Subject Age Group, **Subject ID**, **GO / SC Num**,

Terry Stop ID, **Stop Resolution**, **Weapon Type**,

Officer ID, **Officer YOB**, **Officer Gender**,

Officer Race, **Subject**, **Perceived Race**,

Subject Perceived Gender, **Reported Date**,

Reported Time, **Initial Call Type**,

Final Call Type, **Call Type**, **Officer Squad**,

Arrest Flag, **Frisk Flag**, **Precinct**, **Sector**, **Beat**

VISUALIZATION OF RELATIONSHIP

Arrest Distribution
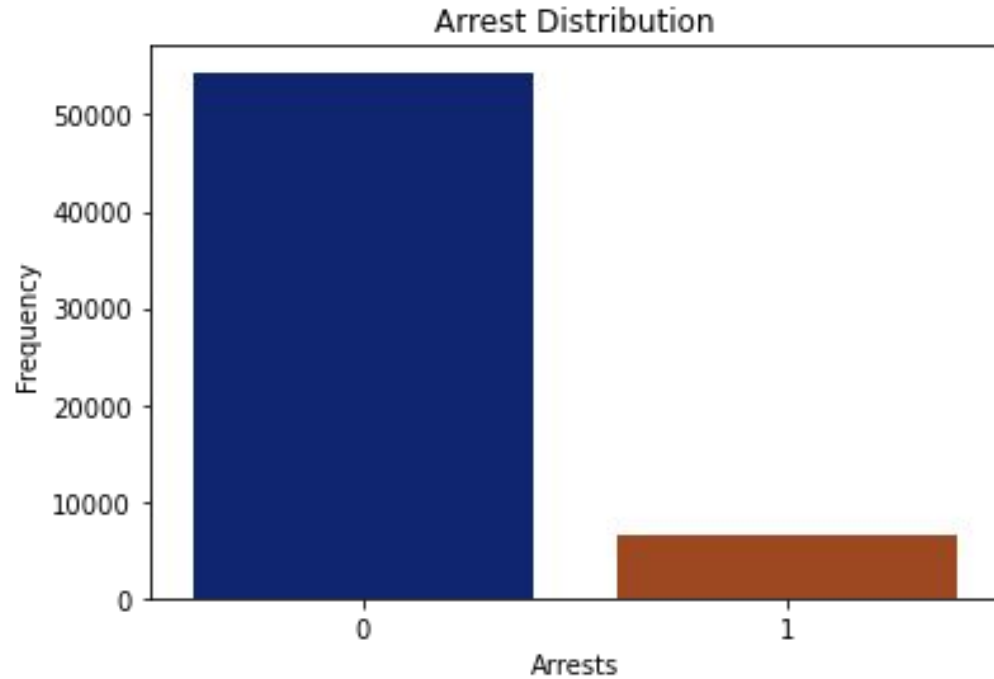
From the above plot the number of arrests made were few after a terry stop

Cont...
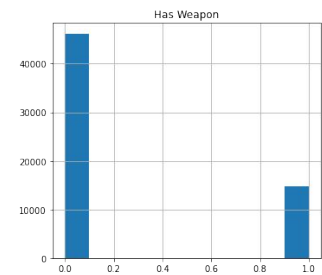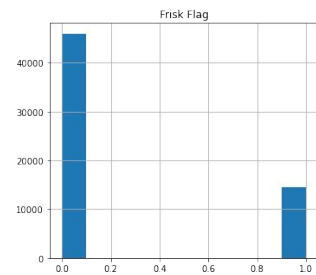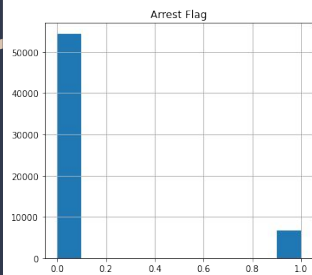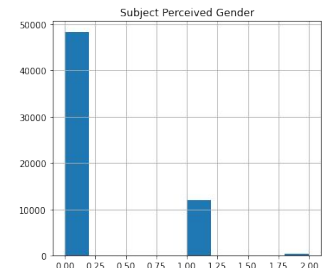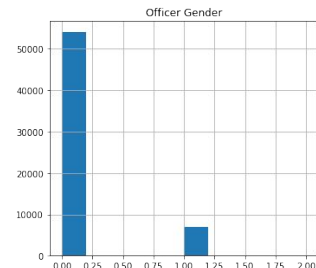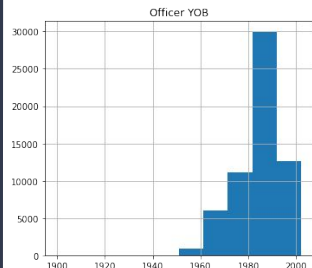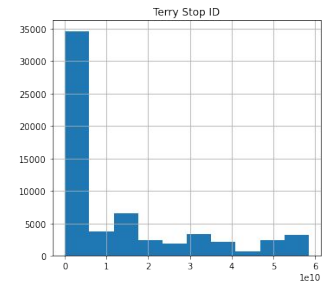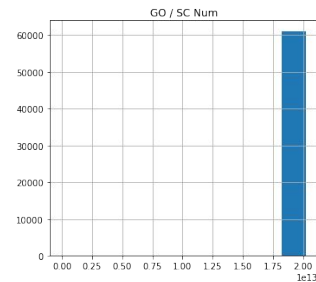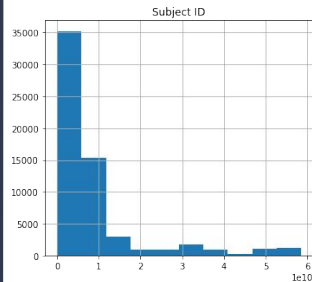
Cont...

Discussion

- Subject ID, GO / SC Num, Terry Stop ID: These graphs show the frequency distribution of different identifiers. The high counts in certain categories might indicate common occurrences or frequent subjects in the dataset.
- Officer YOB, Officer Gender, Subject Perceived Gender: These graphs provide insights into the demographics of officers and subjects. For example, the distribution of officer birth years (YOB) can reveal the age range of the workforce, while gender distributions can highlight diversity or lack thereof.
- Arrest Flag, Frisk Flag, Weapon: These graphs show the frequency of arrests, frisks, and weapon findings. High counts in these categories could indicate areas with more law enforcement activity or higher crime rates.

BI-VARIATE ANALYSIS

CONT…

- Subject Age Group: This plot shows how arrests vary across different age groups. It can help identify which age groups are more frequently involved in Terry Stops leading to arrests.
- Stop Resolution: This plot illustrates the outcomes of Terry Stops, such as whether they resulted in an arrest or another resolution. It can highlight the effectiveness or consequences of these stops.
- Officer ID: This plot shows the distribution of arrests across different officers. It can reveal if certain officers are more likely to make arrests during Terry Stops.
- Subject Perceived Race: This plot examines the relationship between the perceived race of subjects and the number of arrests. It can be used to analyze potential racial disparities in arrests.
- Precinct: This plot shows how arrests are distributed across different precincts. It can help identify areas with higher or lower arrest rates.

EDA CONT...

Multivariate analysis

In this section, we will perform multivariate analysis to examine the relationship between the target variable - price and multiple features in the data. We will use heatmap to visualize the correlation matrix of the features and see how they are related to each other and to the price.

A heatmap can show us the strength and direction of the correlation between two variables using different colors and shades. This will help us identify the most important features for the prediction and avoid multicollinearity problems.
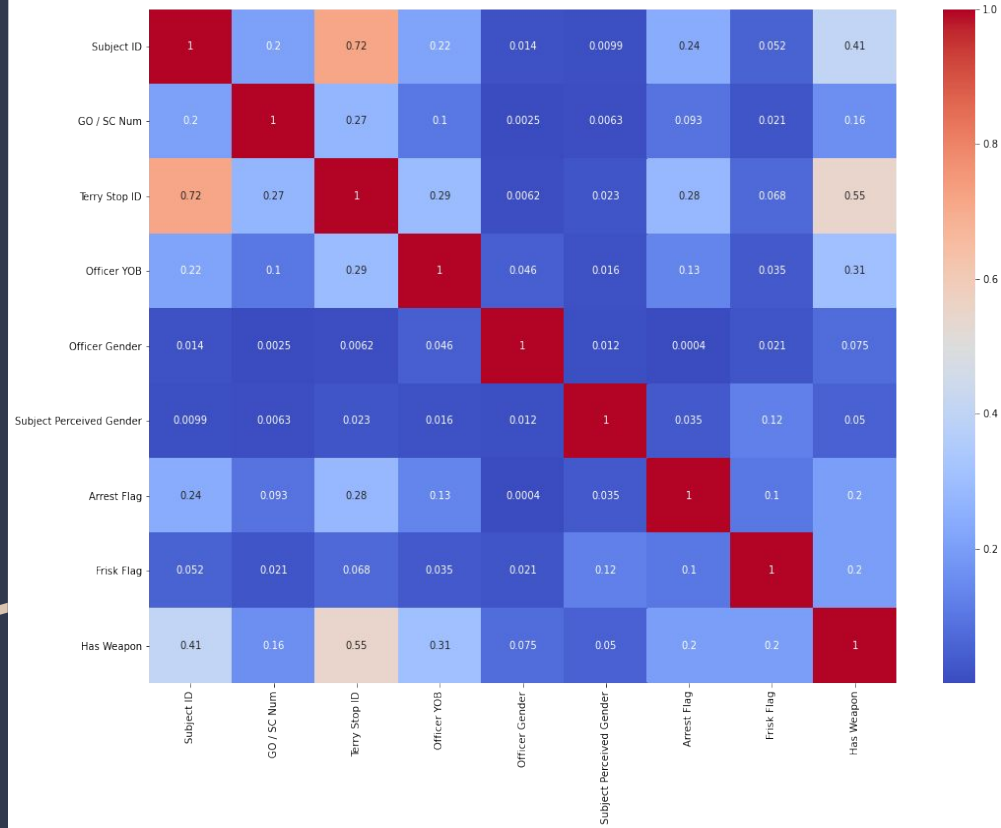
CONT...

Discussion

- Correlation Coefficients: The matrix shows correlation coefficients between variables, ranging from -1 to 1. Positive values indicate a direct relationship, while negative values indicate an inverse relationship.
- Strong Correlations: There are strong positive correlations between variables like 'Arrest Flag' and 'Frisk Flag,' suggesting that frisks often lead to arrests.
- Weak Correlations: Most variables show weak correlations, indicated by values close to zero. This suggests that many factors in Terry Stops are not strongly related to each other.
- Color Coding: The colors range from dark red (strong positive correlation) to dark blue (strong negative correlation), with lighter shades indicating weaker correlations.

MODELING

Processes performed before modelling were:

1. Transforming of Categorical data using Ohe
2. Data splitting using train_test_split
3. Data Scaling using StandardScaler
4. Correcting class imbalance using SMOTE

Baseline Logistic Regression Model

- Precision: The precision values for class 0 and class 1 are 0.939194 and 0.234691, respectively. Higher precision signifies a lower rate of false positives for that class. With a higher precision in class 0, the model demonstrates better performance in predicting class 0 than class 1.

- Recall: The recall values for class 0 and class 1 are 0.764766 and 0.592988, respectively. Recall measures the model's ability to correctly identify positive instances. Like precision, recall is higher for class 0, indicating better performance in identifying class 0 instances compared to class 1.

CONT...

- F1-Score: The F1-scores for class 0 and class 1 are 0.843052 and 0.336287, respectively. The F1-score is the harmonic mean of precision and recall, balancing both metrics. Once again, class 0 has a higher F1-score than class 1.
- Accuracy: The model's accuracy is 0.746135, meaning that 74.61% of the predictions were correct out of all instances.

Therefore, logistic regression achieves 74.61% prediction accuracy on the test data.
Based on these metrics, it is evident that the model performs better for class 0 compared to class 1.

CONFUSION MATRIX FOR LOGISTIC REGRESSION
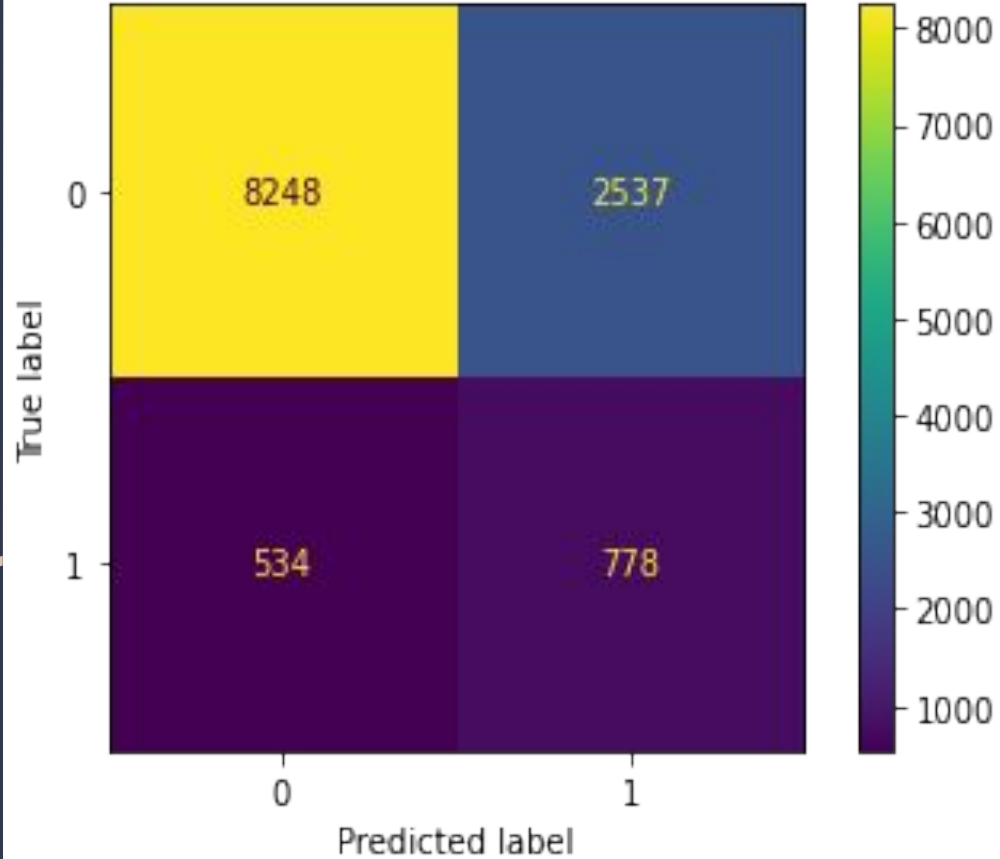
CONT...

Model Performance Summary

The logistic regression model demonstrates the following performance metrics:

- ROC AUC:

Training: 0.6798585343228201

Test: 0.6788768417065255

- Confusion Matrix:

True Positives (TP): 778

False Negatives (FN): 534

True Negatives (TN): 8248

False Positives (FP): 2537

- Accuracy:

Training: 0.6798585343228201

Testing: 0.7461354054724312

CONT...

Observations:

- The model shows a high level of discrimination on the training data but significantly lower on the test data, suggesting potential overfitting.
- The confusion matrix indicates a considerable number of false positives, highlighting the need for model improvement.

CROSS VALIDATION

- After applying cross-validation with five folds, we obtained an improved model with an average accuracy of 0.7461, representing an 74.61% accuracy level in predicting arrests occuring after a Terry Stop on the test data. The training data accuracy also adjusted to 0.7461.
- This adjusted model shows no significant change in performance, accurately predicting the class labels for most instances in both training and testing datasets. The testing accuracy is slightly lower than the training accuracy, which is expected but not significantly different.
- Therefore, we can proceed to evaluate the second model to determine if it performs better in predicting customer churn compared to logistic regression.
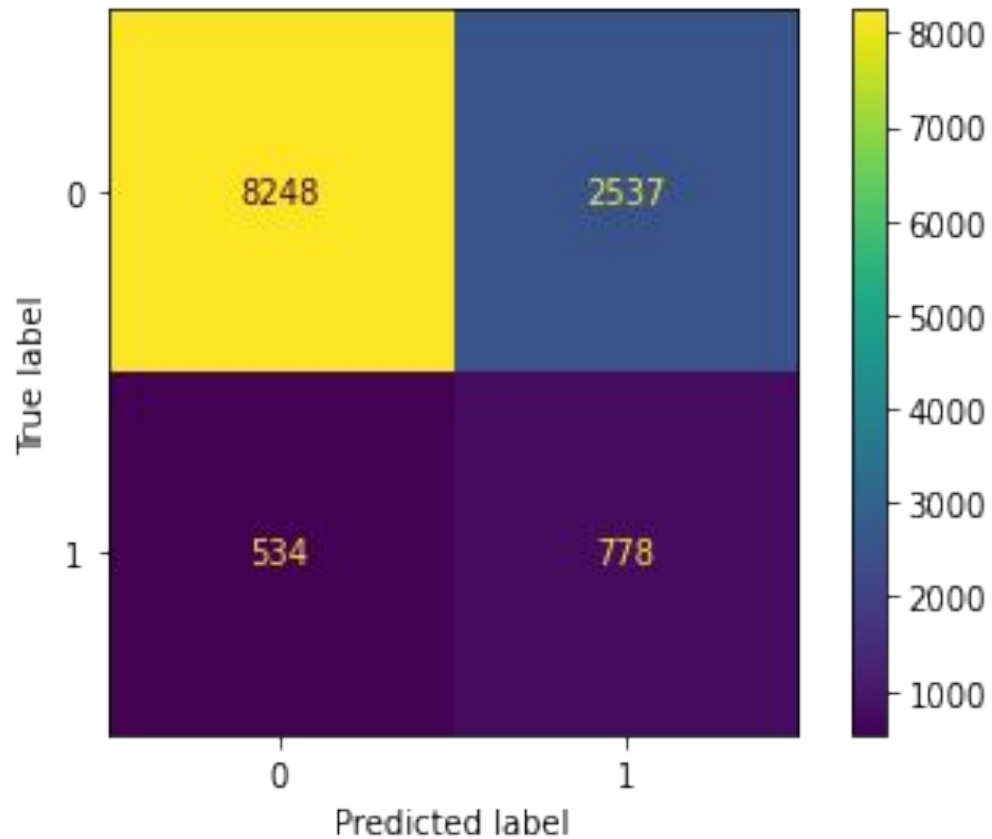
CONT...

Confusion matrix for logistic regression after cross validation.

DECISION TREE CLASSIFIER
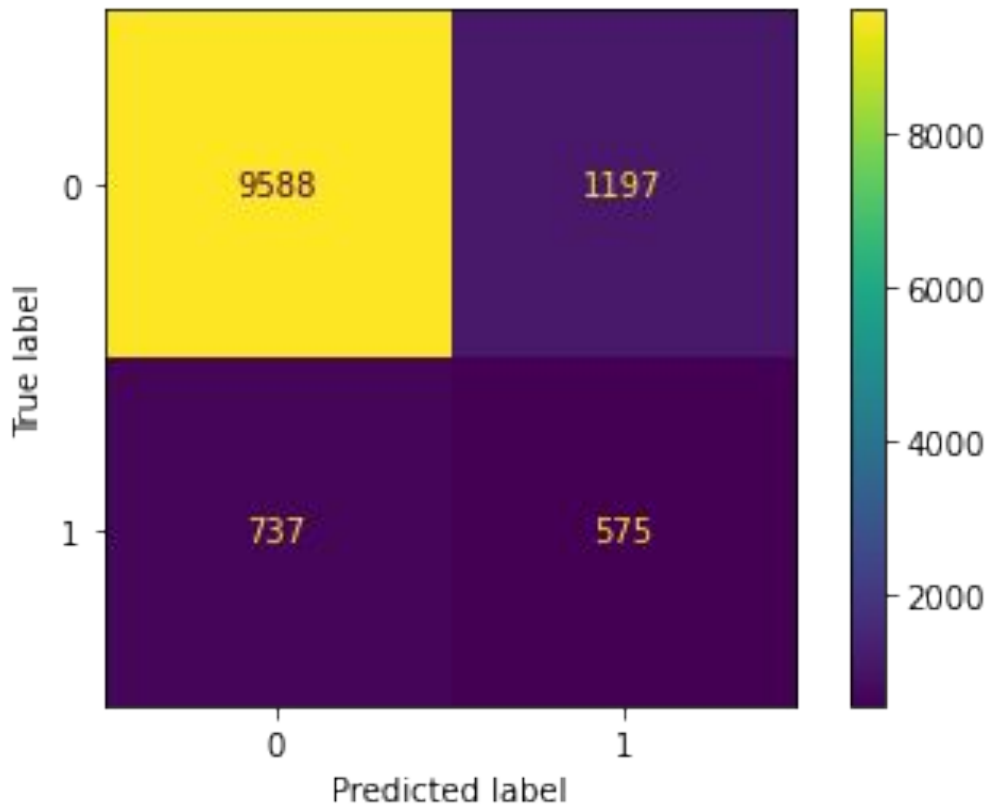
Confusion matrix of a decision tree classifier.

CONT...

The logistic regression model exhibits the following performance metrics:

- Precision:

Class 0: 0.93

Class 1: 0.32

- Recall:

Class 0: 0.89

Class 1: 0.44

- F1-score:

Class 0: 0.91

Class 1: 0.37

- Accuracy:

Overall: 0.84

- Macro Average:

Precision: 0.625

Recall: 0.665

F1-score: 0.64

CONT...

Observations:

- High precision and recall for class 0, indicating strong performance in identifying true negatives.
- The precision is lower in class 1, suggesting the model struggles to correctly identify true positives but the recall is moderate.
- Overall accuracy of 84%, suggesting robust performance across all instances.
- The macro average shows that there is a disparity between the performance on Class 0 and Class 1, with Class 1 being harder for the model to predict correctly.
- The ROC AUC for the test set (0.66) suggests that the model has moderate discrimination ability on unseen data, although there is a notable drop from the training set's ROC AUC (0.88).

To gain further insights, we will examine the classification report for all metrics related to the Decision Tree classifier.

USING RANDOM FOREST TO IMPROVE DECISION TREE MODEL
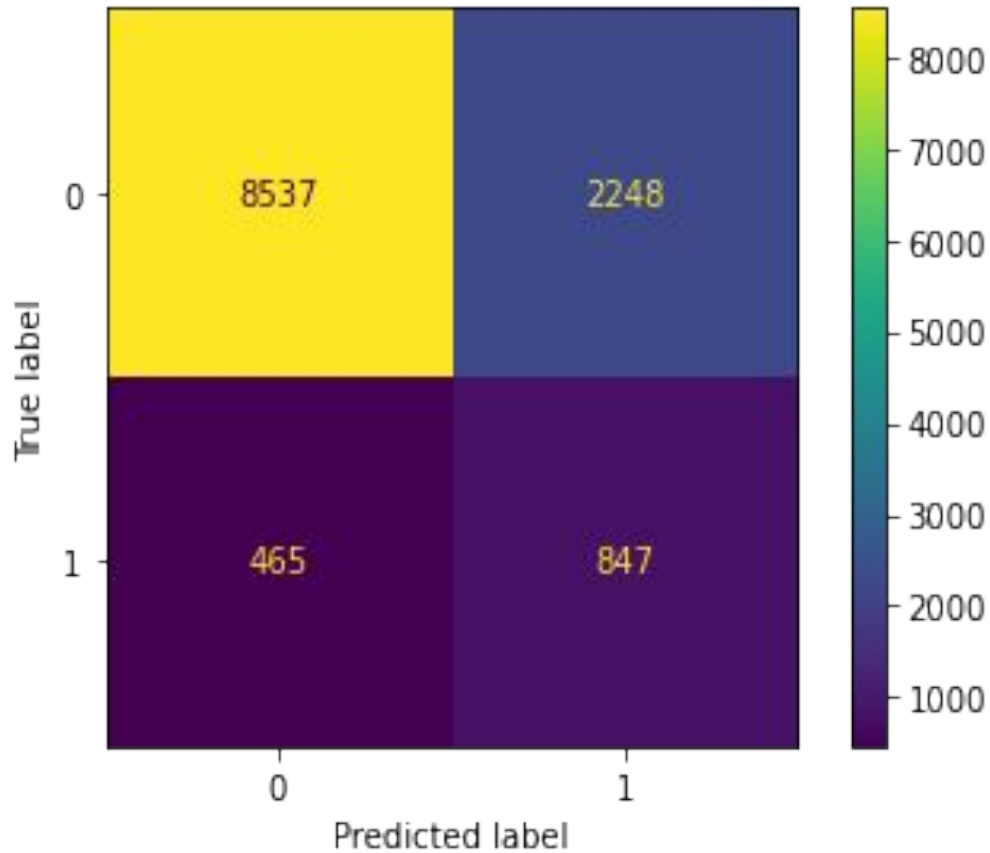
Confusion matrix showing improved decision tree model

CONT...

Confusion Matrix Breakdown

True Positives (TP): 8537
- These are instances where the model correctly predicted the positive class.

False Negatives (FN): 2248
- These are instances where the model incorrectly predicted the negative class when it was actually positive.

False Positives (FP): 465
- These are instances where the model incorrectly predicted the positive class when it was actually negative.

True Negatives (TN): 847
- These are instances where the model correctly predicted the negative class.

CONT...

- High True Positives: The model is quite effective at identifying the positive class, as indicated by the high number of true positives.
- False Negatives: There is a significant number of false negatives, suggesting that the model misses a considerable portion of the positive class.
- False Positives: The number of false positives is relatively low, indicating that the model is less likely to incorrectly identify the negative class as positive.
- True Negatives: The true negatives are also relatively low, which might indicate an imbalance in the dataset or a need for model improvement.

CONT...

Summary

- Based on the analysis of the three models, it is evident that logistic regression performs poorly in predicting arrests after a Terry Stop.
- In contrast, both the Random Forest classifier and Decision Trees perform well with accuracies of 77.6% and 84%, respectively.
- Therefore, it is crucial to enhance the performance of the Random Forest classifier and Decision Trees by tuning hyperparameters to achieve even better accuracy.
- Hyperparameters are essential for improving the efficiency and performance of models.
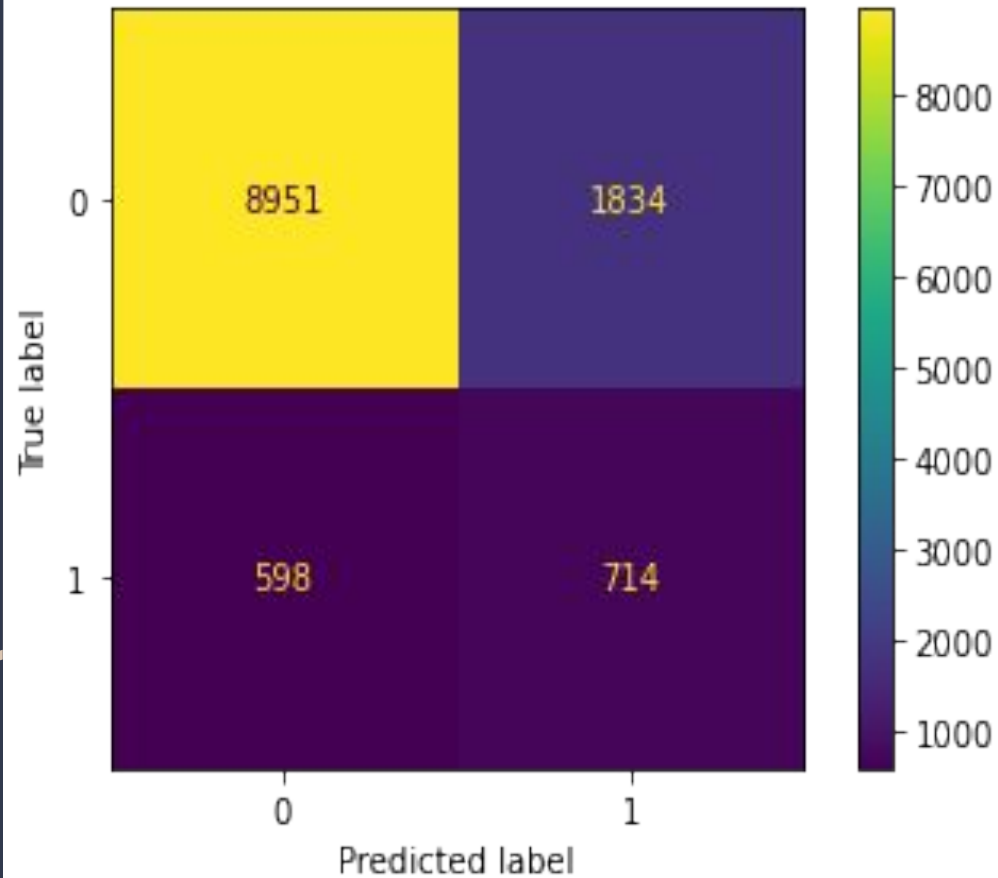
CONT...

Hyperparameter Tuning

- Hyperparameters are settings that are predetermined and not learned from the data during model training.
- They govern aspects like model complexity, regularization, learning rate, and number of iterations.
- As the Random Forest classifier is our best-performing model among the three, we will employ Gridsearch to enhance its performance.

RANDOM FOREST MODEL

- After optimizing the parameters for the Random Forest classifier using grid search, our model showed improved performance as described below:

- The accuracy of the Random Forest model now stands at 79%, indicating that it correctly predicted the class labels for the test data with around 79.9% accuracy and achieved 88.7% accuracy on the training data, making it highly effective for predicting arrests occurring after a Terry Stop.

- The confusion matrix with tuned parameters shows 714 true positives (TP), 8951 true negatives (TN), 1834 false positives (FP), and 598 false negatives (FN), demonstrating excellent predictive capability.

CONFUSION MATRIX FOR RANDOM FOREST MODEL

TUNED RANDOM FOREST MODEL CONFUSION MATRIX
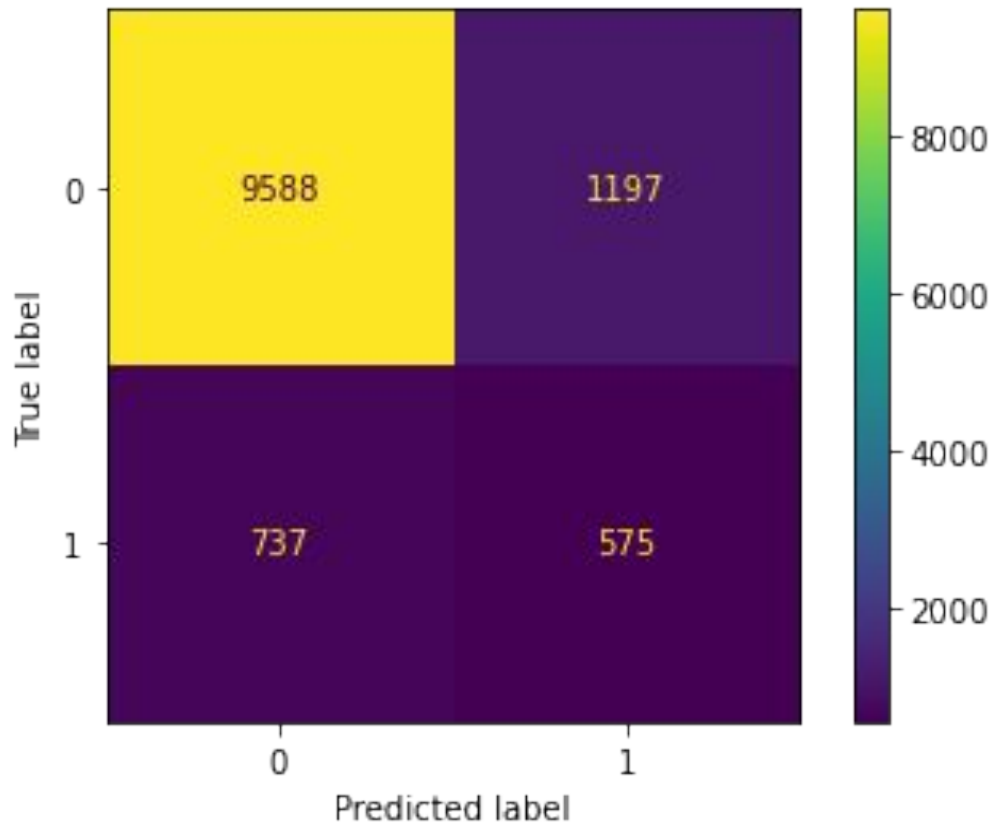
CONT…

- The tuned Decision Tree model achieved a training accuracy of 87.5% and an improved test accuracy of 88%. This suggests that the model is performing well and has learned patterns in the data that enable accurate predictions.
- The Logistic Regression model attained an AUC-ROC score of around 0.6799 on the training data and 0.6789 on the test data.

EVALUATION

- We have developed three machine learning models to predict arrests occurring after Terry Stops in Seattle. Upon testing, we found that logistic regression performs poorly, achieving a training accuracy of 74.61% and a testing accuracy of 67.98%. Despite applying 3-fold cross-validation to mitigate overfitting, the training accuracy dropped to 67.99% and the testing accuracy to 67.89%.

- In contrast, the decision tree classifier and random forest models demonstrated better accuracy. The decision tree achieved training and testing accuracies of 87.99% and 84.01%, respectively, while the random forest model had a training accuracy of 87.47% and a testing accuracy of 77.57%.

CONT...

- Therefore, it is evident that the decision tree model had the best average prediction accuracy, followed by the random forest. To further improve our prediction accuracy, we applied hyperparameter tuning using GridSearch.

- After hyperparameter tuning, the decision tree model's accuracy improved to 88.7% on the training data and 79.9% on the testing data. The random forest model's accuracy improved to 88.7% on the training data and 79.9% on the testing data.

- This indicates that there was some overfitting in the random forest model, making the decision tree preferable. In terms of precision, recall, and F1 score, the tuned decision tree outperformed the random forest classifier.

ROC CURVES

Receiver Operating Characteristic (ROC) Curves

- Random Forest (AUC = 0.85)
- Logistic Regression (AUC = 0.80)
- Decision Tree (AUC = 0.84)
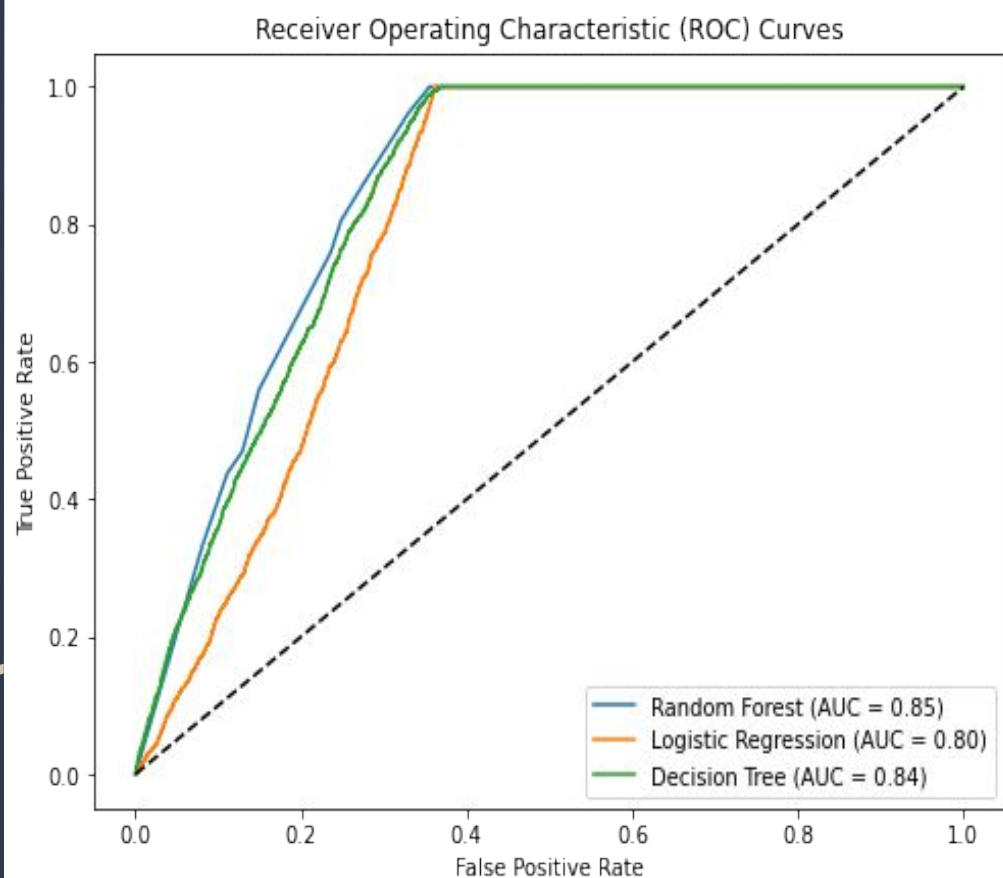
MODEL COMPARISON SUMMARY

Based on the AUC (Area Under the Curve) values from the ROC (Receiver Operating Characteristic) curve, the following inferences can be made about the models' performance:

- **Decision Tree:**

AUC: 0.84

Inference: Exhibits the strongest discriminatory power, effectively distinguishing between positive and negative classes. Maintains a high true positive rate (TPR) with a low false positive rate (FPR), resulting in a large ROC curve area.

- **Random Forest:**

AUC: 0.85

Inference: Demonstrates good classification ability slightly better than the Decision Tree. May have marginally lower false positive and false negative rates.

CONT…

- Logistic Regression:

AUC: 0.80

Inference: Shows the weakest discriminatory performance among the three models, with potentially higher false positive and false negative rates.

CONT...

Conclusion:

- The Random Forest is identified as the best model for Arrests predictions due to its largest area under the ROC curve and highest accuracy and the most improved hence best in predicting Arrest occurring after Terry Stops.
- The Decision Tree also performs well but is slightly less effective than the Random Forest.
- The Logistic Regression model, while useful, has the lowest AUC, indicating comparatively weaker performance in distinguishing between classes.

CONCLUSION

By leveraging the best model, which is the Random Forest, the public safety department can achieve significant benefits:

- Accurate Prediction of arrests after a Terry stop: The model's high accuracy ensures effective identification of arrests likely to occur, enabling preparation on how to prosecute such cases.
- Resource Allocation: Efficient Use of Resources: By predicting the likelihood of an arrest following a Terry stop, departments can allocate resources more effectively. This ensures that officers are deployed where they are most needed, potentially reducing response times and improving overall efficiency.
- Crime Prevention through proactive Policing: Predictive models can help identify patterns and hotspots for criminal activity. This allows officers to be more proactive in their patrols, potentially preventing crimes before they occur.

CONT...

- Improved Training and Policies through data-driven decisions: Analysis of predictive data can inform training programs and policy adjustments. Departments can identify areas where officers may need additional training or where policies might need to be revised to improve outcomes.
- Community Trust (Transparency and Accountability): Using data to predict arrests can increase transparency in policing practices. When the public sees that decisions are based on data and not bias, it can help build trust between the community and the police.

CONT...

- Legal and Ethical Compliance reducing bias: Predictive models can help ensure that Terry stops are conducted based on objective criteria rather than subjective judgment, potentially reducing instances of biased policing.
- Enhanced Officer Safety: Predicting the likelihood of an arrest can also help assess the risk level of a situation, allowing officers to take necessary precautions to ensure their safety.

THANK YOU !!

FOR CONTACTS:
EMAIL: bennie.kibet@student.moringaschool.com
Linked in: Bennie Tonui