Terry Stops:

Predicting Outcome

Outline

- Overview
- Business Understanding
- Data Understanding

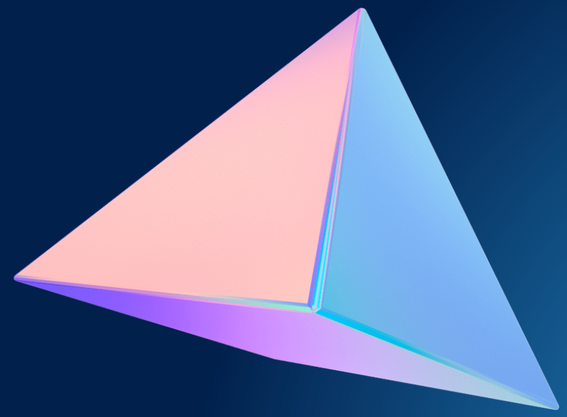
- Modelling
- Evaluation

- Recommendations
- Next Steps

Overview

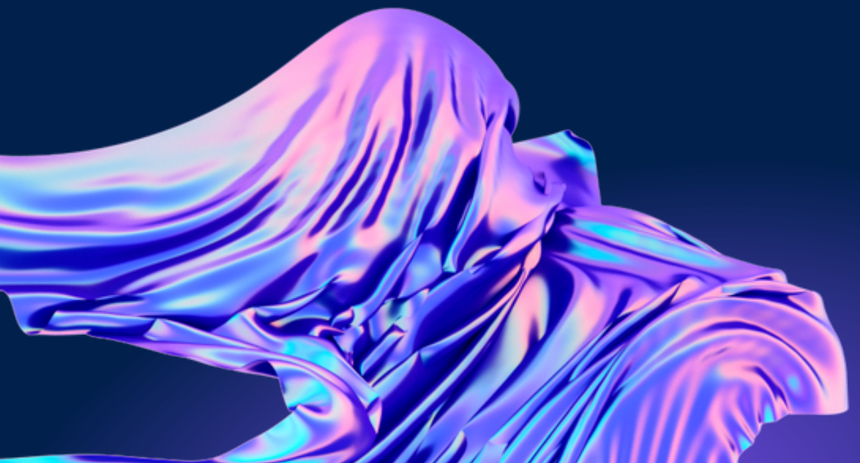
Terry V. Ohio set the precedent that a police officer can stop you for looking suspicious. This led to it being called a Terry Stop.

Problem Statement



Build a model that will accurately predict the outcome of a Terry Stop.

i.e Arrest or No Arrest



Data Understanding

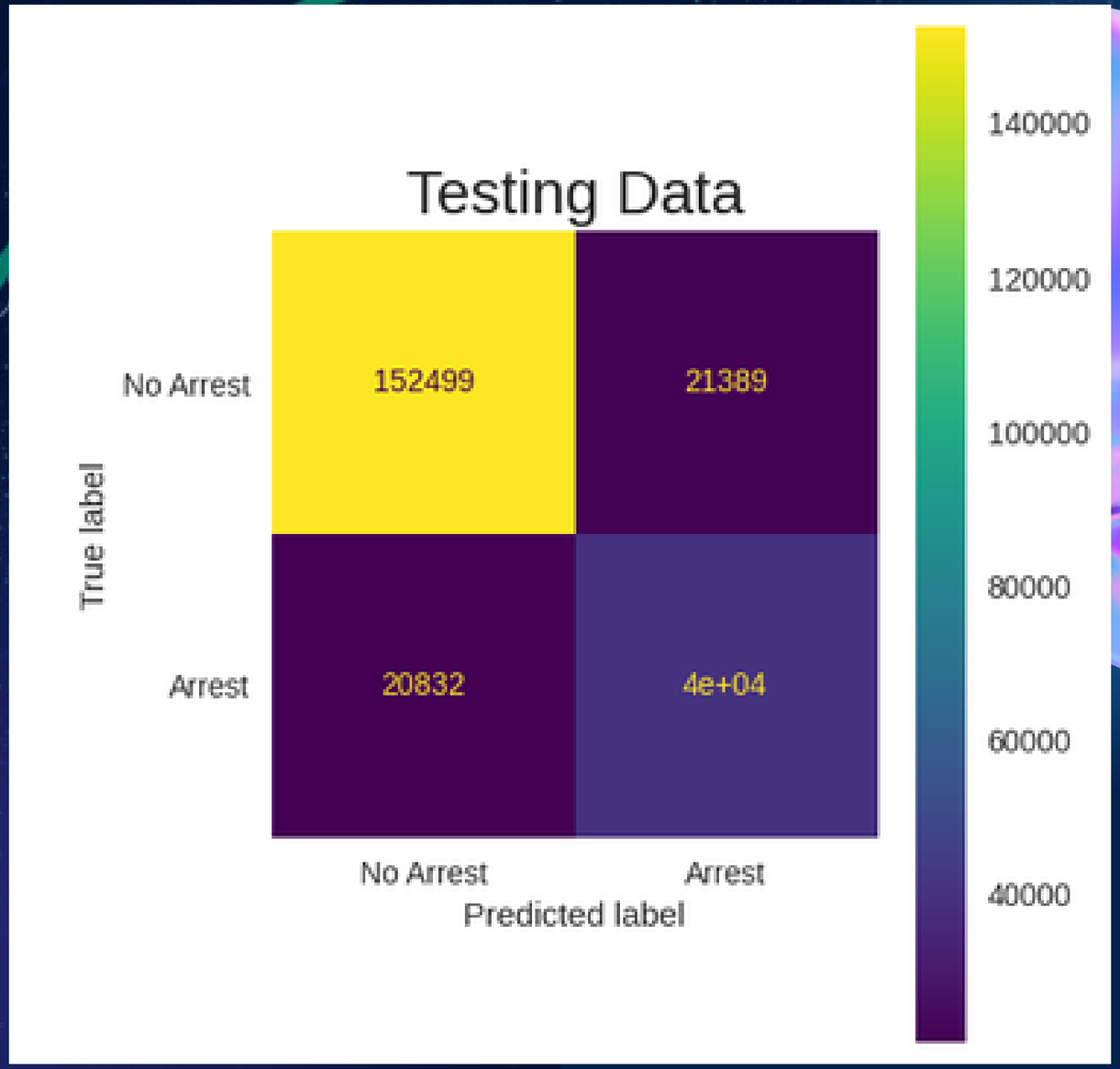
- Contains attributes related to the stops.
- Had 2.62 Million rows by 23 columns.
- 38.8% in one column was missing.
- Other columns had placeholders.

Modelling

- There were six models developed.
- One model was a tuning of a previous model but produced similar numbers.
- Used Stop Resolution as the target.
- Some of the models produced results that did not meet the threshold.
- Other models were computationally taxing and their runtime was too long.

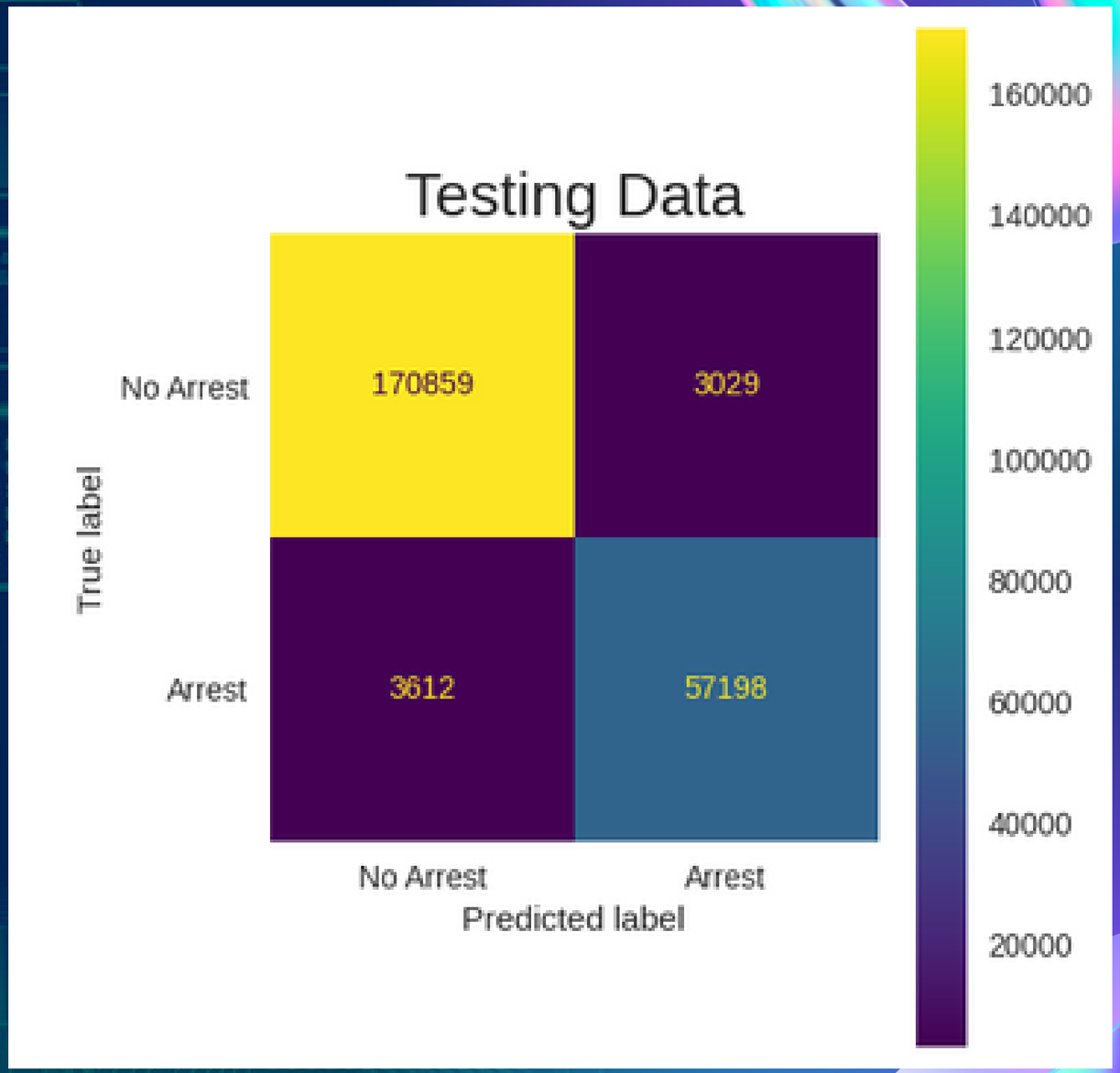
Logistic Regression

- It had a Precision score of 0.65.
- This means that it correctly identified 65% of the outcomes of a stop.
- It managed to identify 152, 499 correctly as No Arrest and 40, 000 as Arrest.
- It still had very high numbers of misidentification
- This is very low for our target.



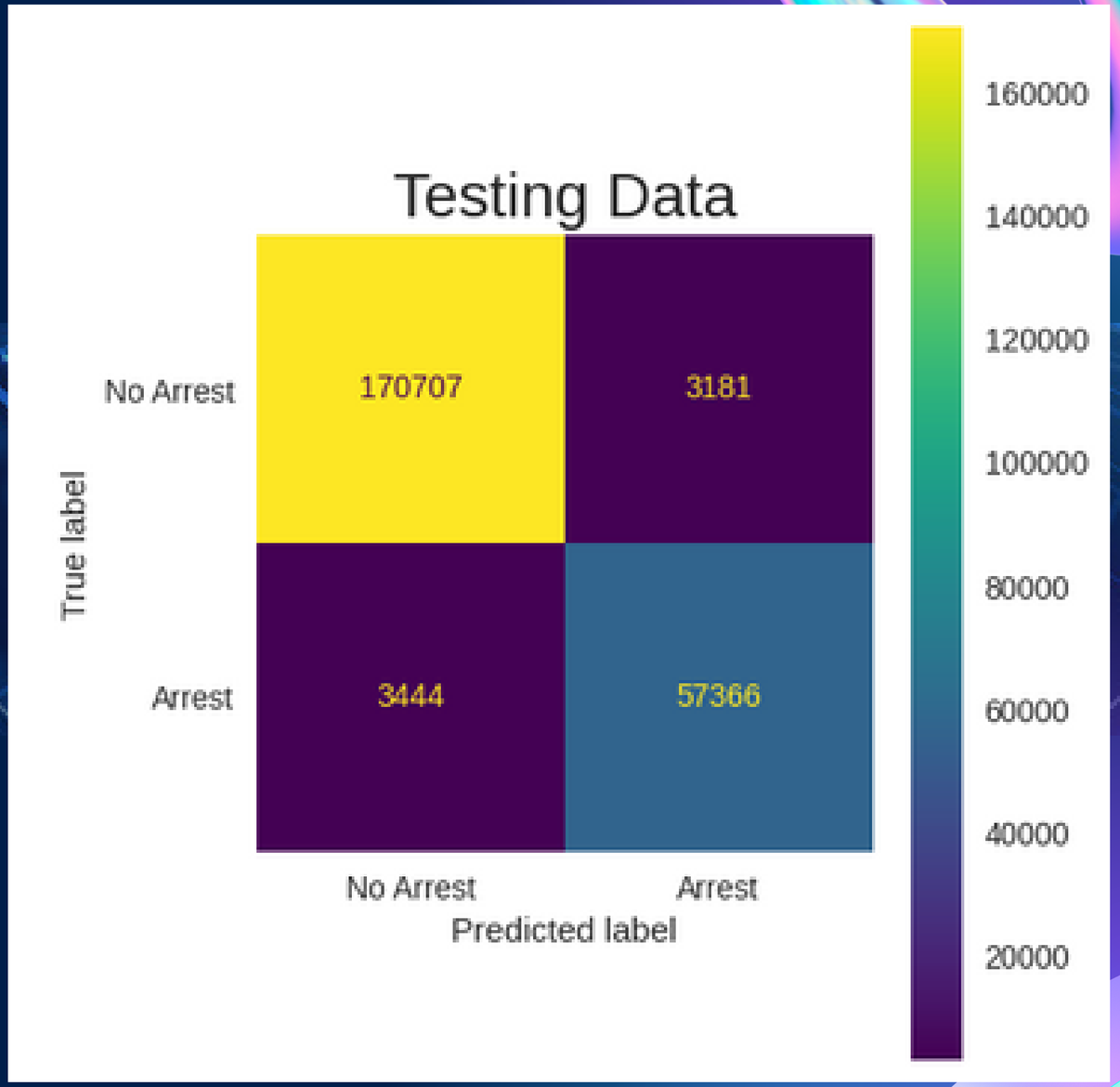
Decision Tree

- It got a 0.95 precision score.
- This indicates that it had a 95% accuracy in identifying Arrests or No Arrests.
- It identified 170, 859 as No Arrest and 57, 198 as Arrest correctly.



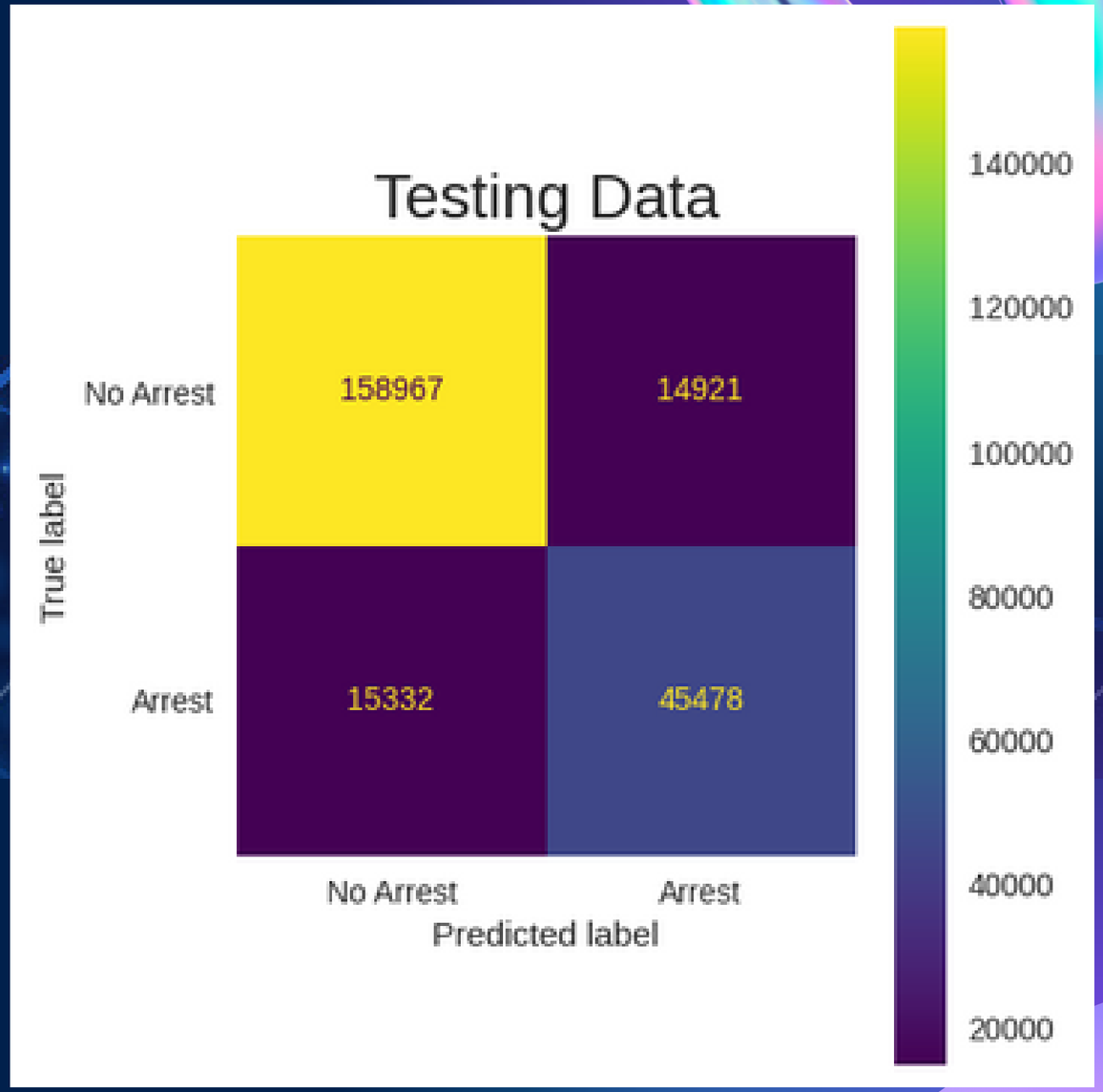
Random Forest

- It was able to correctly identify 95% of the outcomes.
- However it is time consuming and requires a lot of machine power.
- It identified 170707 as No Arrest and 57366 as Arrest correctly.



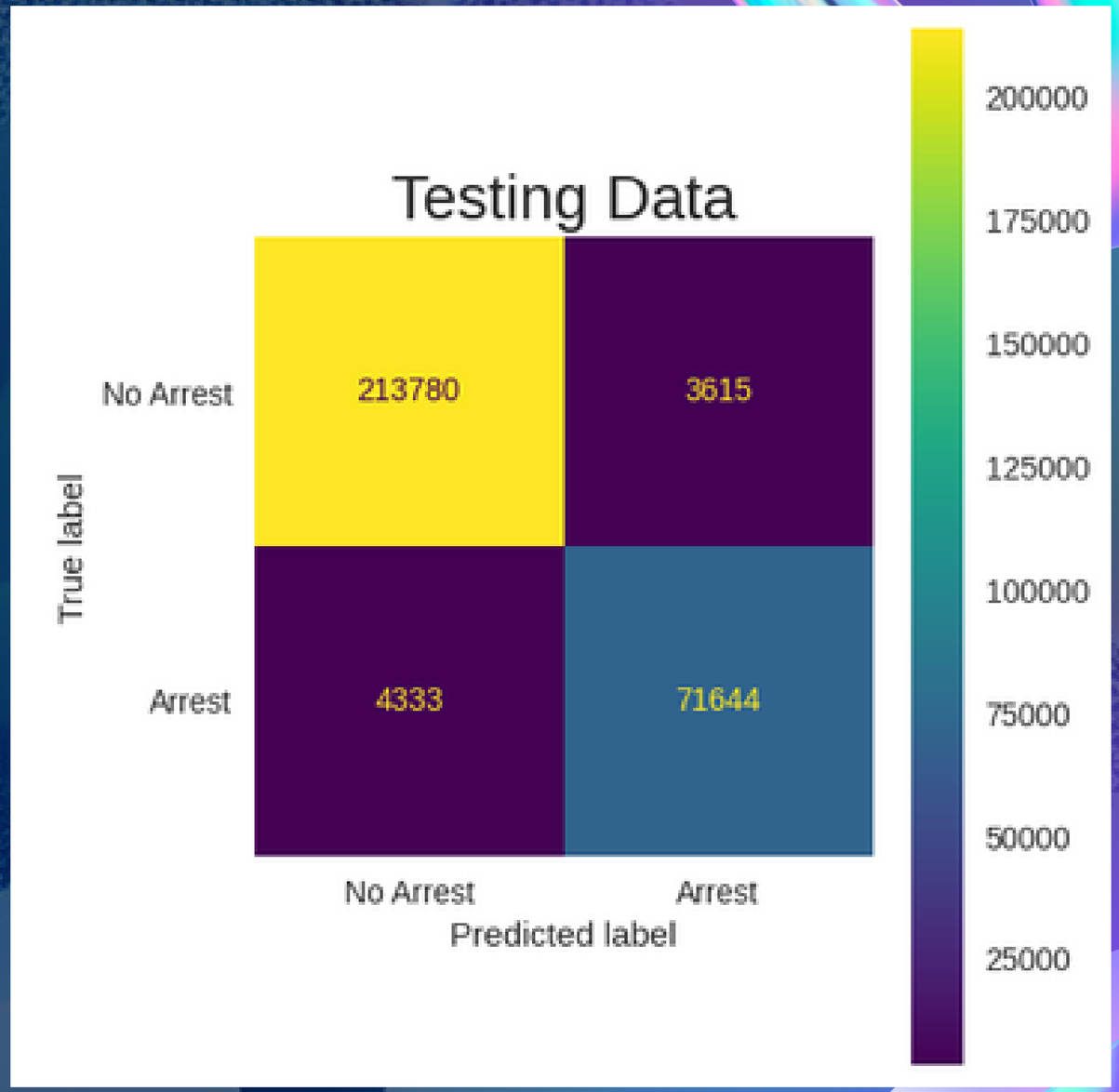
XGBoost

- Attained a precision score of 0.75.
- To our target, a 75% accuracy is not acceptable.
- It identified 158, 967 as No Arrest and 45,478 as Arrest accurately.
- The numbers predicted wrong are too high for our target.



Final Model: Decision Tree

- Attained a precision score of 0.95.
- This an acceptable accuracy as it could identify the outcome with an accuracy of 95%.
- Using data the model had not seen before, it was able to identify 213,780 as No Arrest and 71, 644 as Arrest accurately.



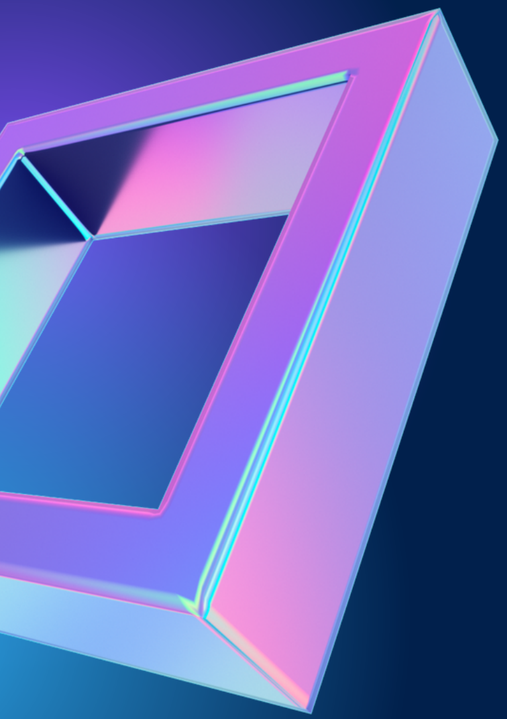
Recommendations

1. Screening Tool: From its accuracy, the model can be used to screen to give priority to cases that may result in arrests. It can help to allocate resources effectively.
2. Identify areas for further investigation: Analyse the predictions made and look at cases where it made incorrect predictions to provide insight into areas that might require further investigation. It can help identify bias, gaps in the data or other factors that maybe influential.
3. Consider ethical implications: The model may suffer from biases such as racial bias. The model should be investigated for any inequalities to different demographics.
4. Combine model predictions with human judgment: The model should only support human decision making but not replace it. It can enhance effectiveness but human judgement should still be at the front.

Conclusions

1. Have multiple outcomes: Instead of grouping all outcomes to either arrest or no arrest, the target can have multiple outcomes.
2. Model data without grouping: Various features had to be grouped together because of computing power but with enough power the model can try to use all features as they are.
3. Analyse effects of grouping: Both features and target were grouped. This may have led to the high accuracies observed. It can be investigated if this is the case.

Thank You



For any questions, reach me at:

GitHub: Robert Mbau

LinkedIn: Robert Mbau

Email: Robert Mbau