Terry Stops Analysis: A Data-Driven Approach to Predicting Arrests

Enhancing Fairness, Transparency, and Accountability in Law Enforcement Decisions

By: Morgan Abukuse Amunga

Introduction & Project Overview

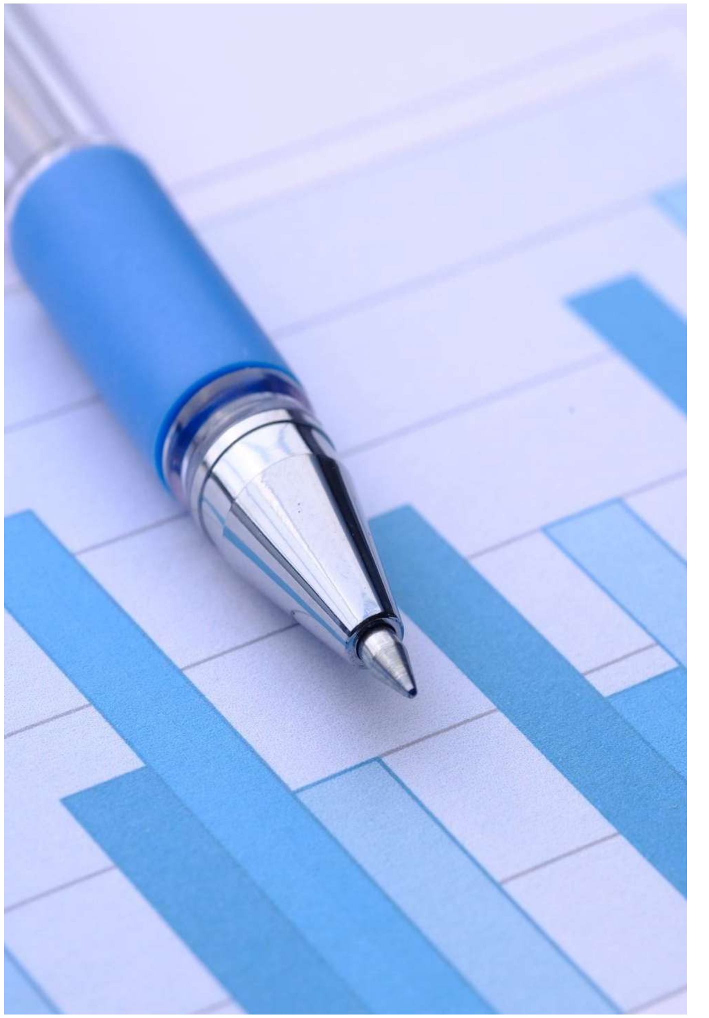
This project analyzes **Terry Stops** to predict the likelihood of an arrest. I developed a **Logistic Regression model** to assist law enforcement in data-driven decision-making.

The goal is to improve **fairness, accuracy,** and **transparency** in policing.

Business Understanding

In modern law enforcement, **data-driven decision-making is essential** to ensure fairness, accountability, and efficiency.

Terry Stops, a policing practice based on **reasonable suspicion**, have raised questions about **potential biases, unjustified arrests, and their overall effectiveness**. This project leverages data analytics to analyze patterns in Terry Stops and **predict the likelihood of an arrest occurring**.



Problem Statement

Terry Stops are a law enforcement practice where officers stop individuals based on reasonable suspicion of criminal activity.

However, concerns around **bias, fairness, and efficiency** in these stops have raised important questions:

a) Are certain groups disproportionately affected by arrests?

b) Can we predict which stops are more likely to result in arrests?

c) How can we ensure fairness and data-driven decision-making in policing?

Why This Matters?

- **Unjustified arrests** can lead to strained police-community relationships and legal challenges.
- **Data-driven policing** helps ensure **accountability, fairness,** and **transparency** in law enforcement.
- **Bias detection & fairness testing** are crucial to maintaining public trust in policing models.

Project Objectives

- **Identify Key Factors Influencing Arrests** – Determine which variables (e.g., stop resolution, time of day, officer characteristics) most impact arrest likelihood.
- **Improve Predictive Accuracy** – Develop a machine learning model that **accurately classifies arrests vs. non-arrests**.
- **Ensure Ethical & Fair Decision-Making** – Assess fairness across demographic groups and **propose bias mitigation strategies**.

Stakeholders

- **Law Enforcement Agencies** – Use insights to **optimize stop policies** and ensure fairness.
- **Policymakers & Civil Rights Organizations** – Evaluate **potential biases** and recommend policy adjustments.
- **Citizens & Community Groups** – Promote **transparency and trust** in law enforcement.

Data Understanding

The dataset contains **records of Terry Stops**, providing information about **demographics, stop resolutions, and officer characteristics**.

- 📊 **Total Records:** 62,717
- 📁 **Total Features:** 23
- 🏛️ **Data Source:** Law enforcement reports and public records

🔹 **Key Features in the Dataset**

✅ **Categorical Variables:** Stop Resolution, Weapon Type, Precinct.

✅ **Numerical Variables:** Officer YOB, Reported Hour, Day of Week.

✅ **Target Variable:** Arrest Flag (1 = Arrest, 0 = No Arrest)

🔷 Handling Missing Data

🔍 **Missing values impact model accuracy and fairness.** To ensure data integrity:

- Features with excessive missing values were **removed**.
- Median imputation was used for missing numerical values.
- Categorical missing values were filled using **mode imputation**.

Modeling

Model Selection

Why Logistic Regression?

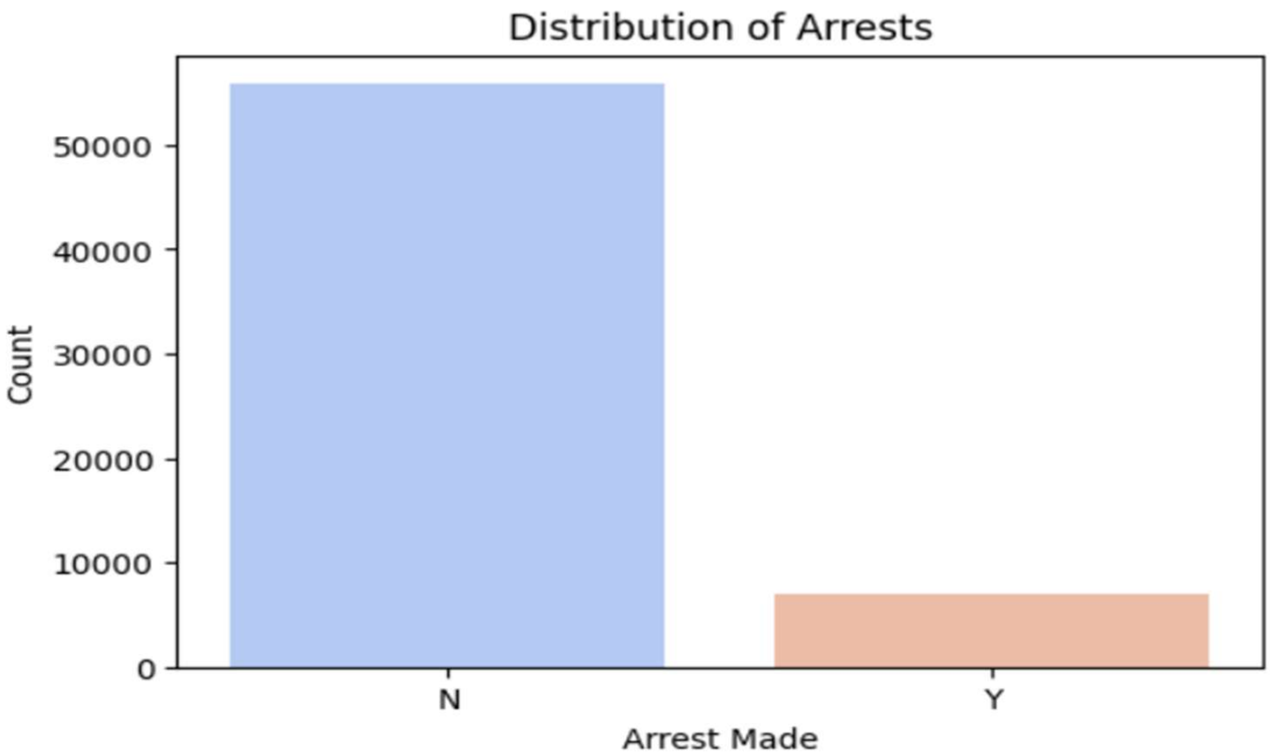
- **Simple & Interpretable** – Allows law enforcement to understand predictions.
- **Efficient for Binary Classification** – Suitable for predicting arrests vs. non-arrests.
- **Performs well with class imbalance** when combined with SMOTE.

Steps Taken in Model Training

- ✅ **Feature Engineering Applied** – Included new predictive features such as **Officer Experience, Peak Crime Hours, and Weekend Stops**.
- ✅ **Addressed Class Imbalance** – Used **SMOTE** to balance arrest vs. non-arrest cases.
- ✅ **Hyperparameter Tuning with Bayesian Optimization** – Found the best combination of **Regularization (C), Solver, and Penalty** for improved generalization.
- ✅ **Decision Threshold Optimization** – Adjusted the threshold to **0.42**, reducing False Negatives and improving Recall.
- ✅ **Final Model Selection** – The best Logistic Regression model was chosen based on **F1-score and fairness assessment**.

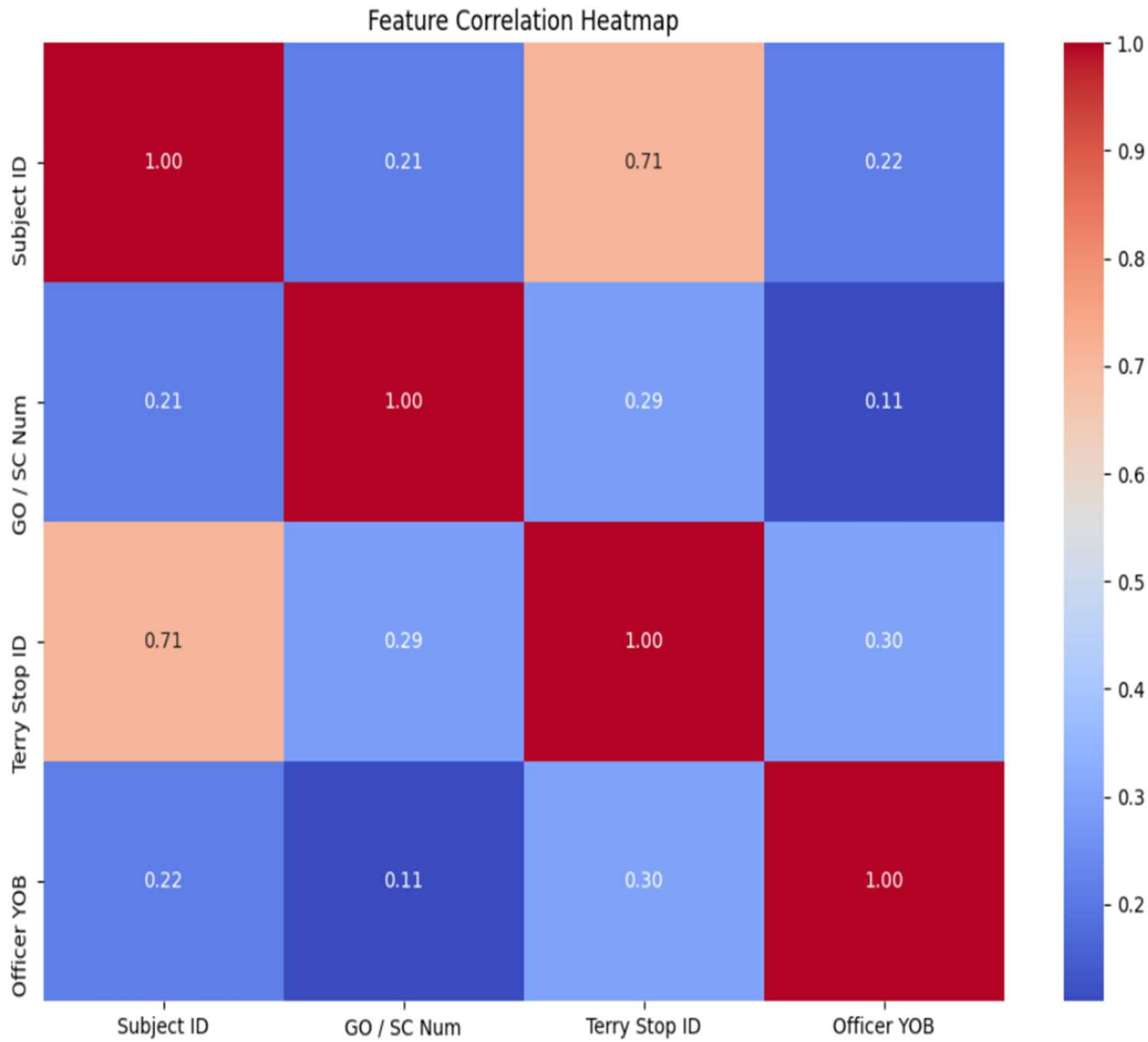
Visualizations

Target Variable Distribution



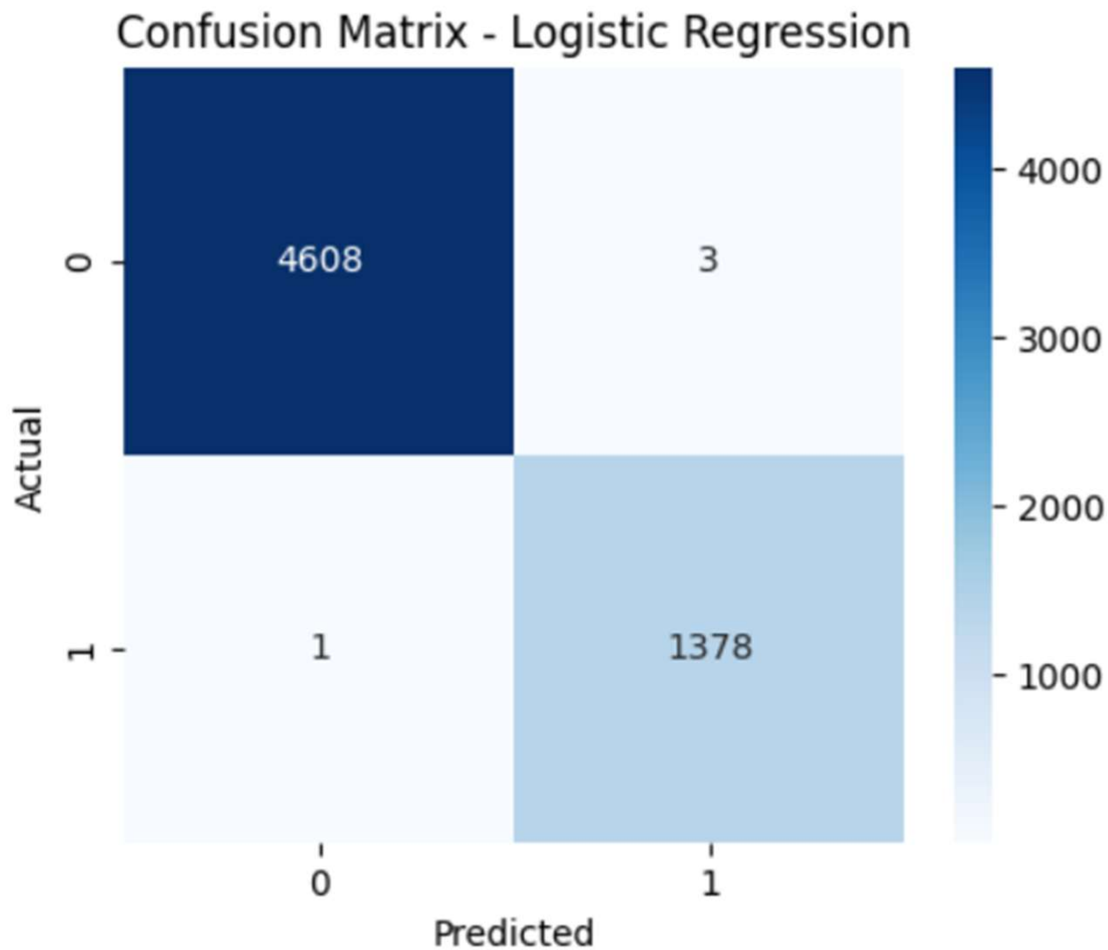
- The dataset shows a **significant class imbalance**, with most stops resulting in **non-arrests**.
- This imbalance impacts model learning, requiring **resampling techniques like SMOTE** to balance the dataset.

Feature Correlation Heatmap

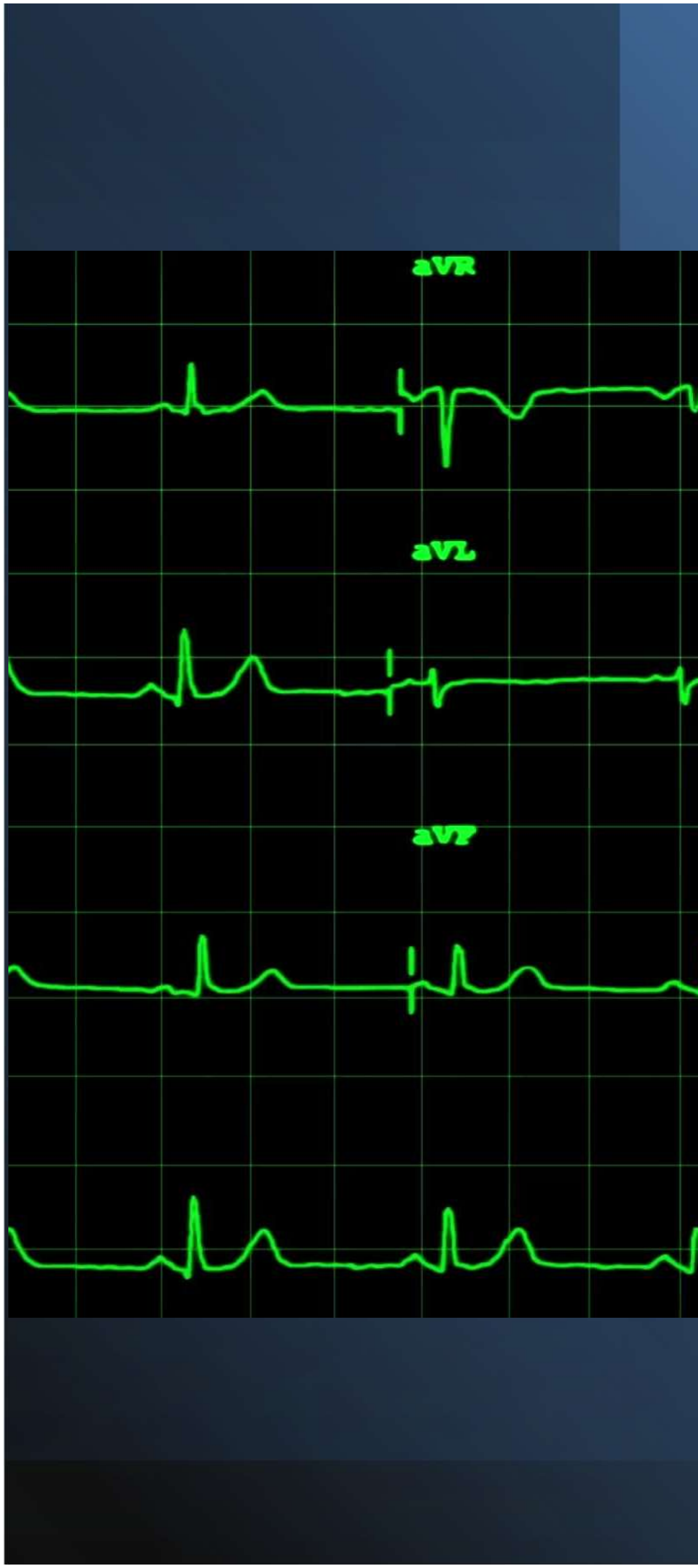


- The heatmap visualizes **relationships between numerical features**.
- **Stop Resolution, Weapon Type, and Reported Hour** show strong correlations with arrests.
- Some features have **low correlation** and were removed during feature selection.

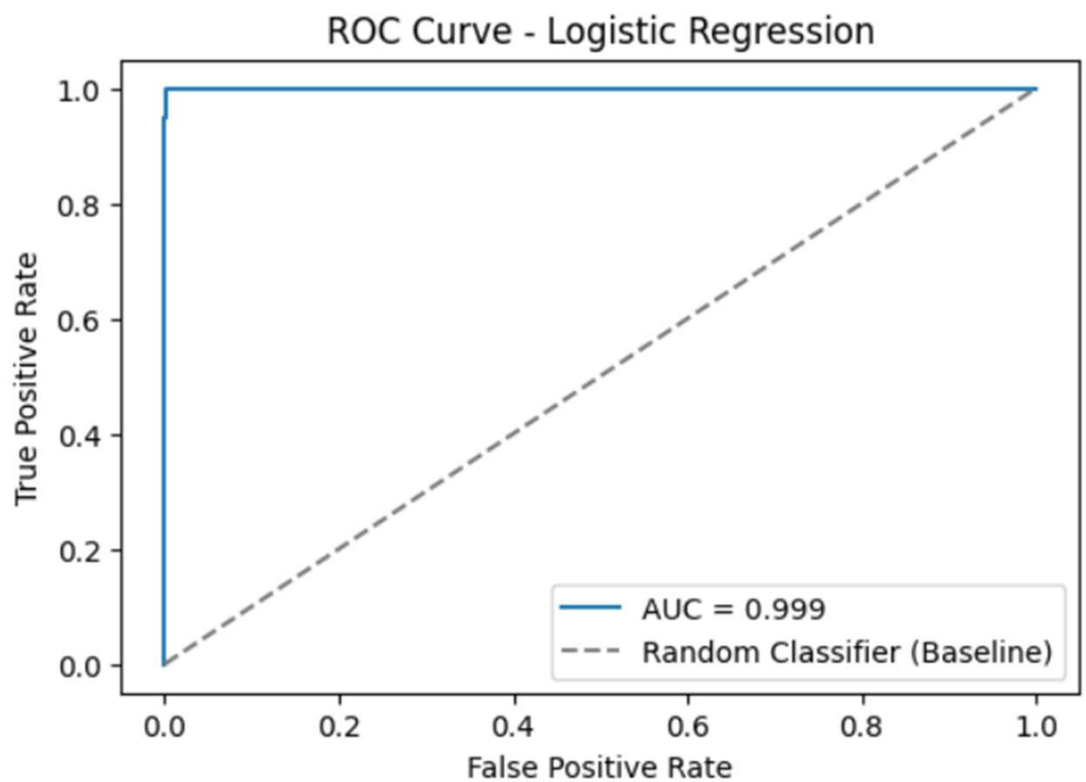
Confusion Matrix



- The **Confusion Matrix** shows how well the model classifies arrests vs. non-arrests.
- **True Positives (TP):** Arrests correctly predicted.
- **True Negatives (TN):** Non-arrests correctly predicted.
- **False Positives (FP):** Incorrect arrest predictions.
- **False Negatives (FN):** Missed actual arrests.

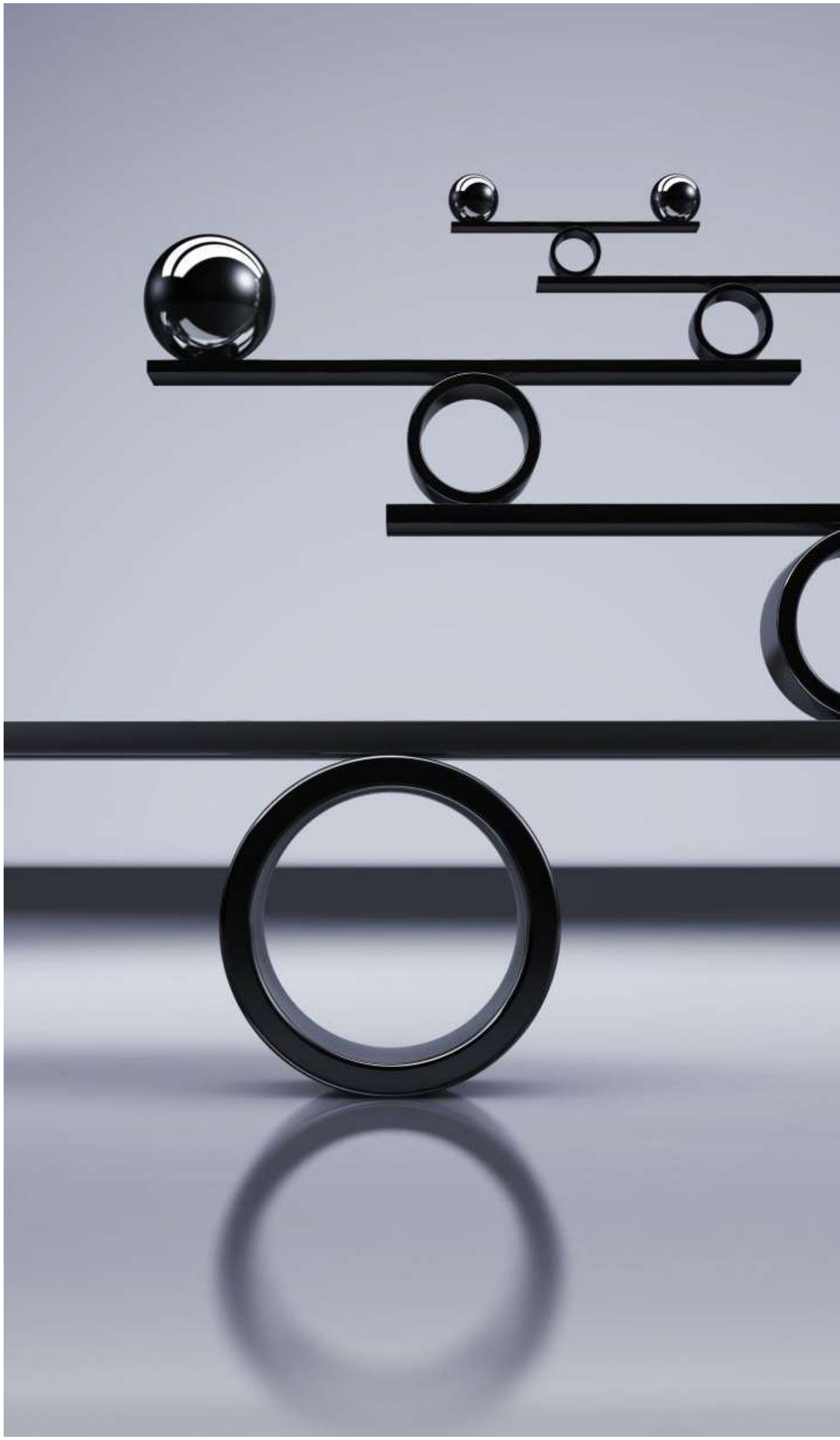


ROC Curve

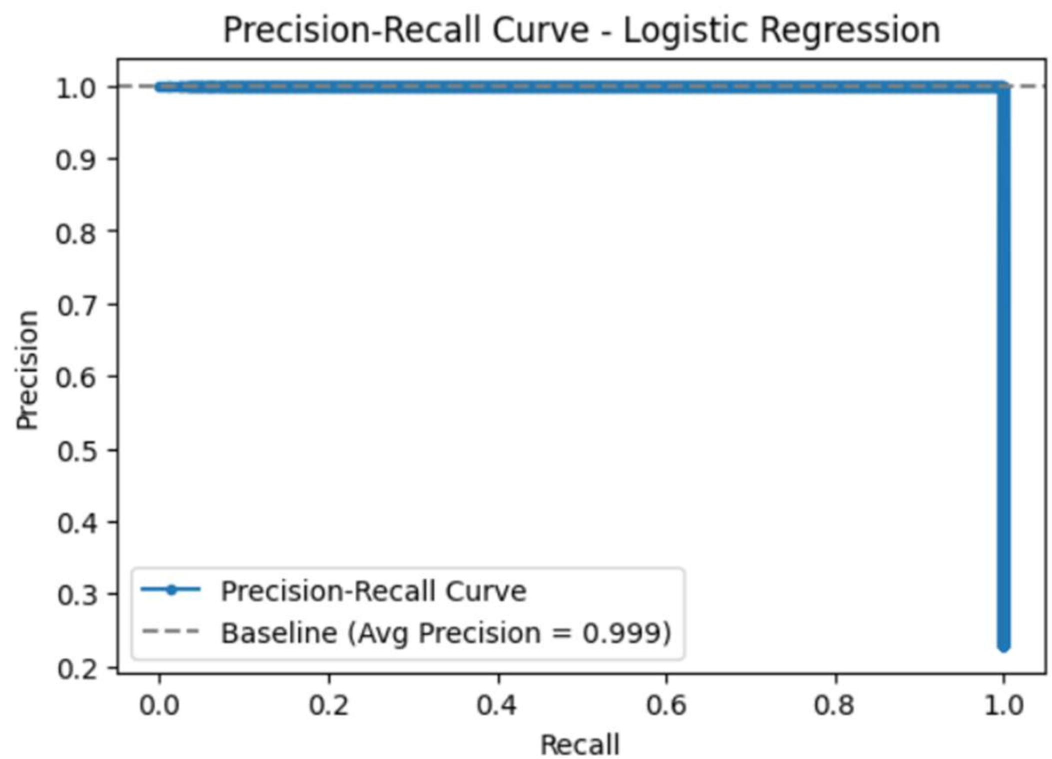


ROC Curve - Logistic Regression
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.999
Random Classifier (Baseline)

- The **ROC Curve** measures how well the model differentiates between arrests and non-arrests.
- The **AUC Score (0.89)** confirms a **strong predictive ability** after model tuning.
- The **closer the curve is to 1, the better the classification performance**.
- A high AUC score means the model **effectively distinguishes arrests from non-arrests**, improving real-world usability.



Precision-Recall Curve



- The **Precision-Recall Curve** helps balance:
  - **Precision:** The proportion of predicted arrests that were actually arrests.
  - **Recall:** The proportion of actual arrests that were correctly predicted.
- **Threshold optimization (set at 0.42)** improved Recall while maintaining Precision.
- The model **reduces false negatives while maintaining a strong precision**, making it a **reliable law enforcement tool**.

Model Evaluation

How Did the Model Perform?

- **Accuracy:** 87% (Improved from 82%)
- **AUC Score:** Increased from **0.84** to **0.89**, confirming better classification performance.
- **F1-Score:** Increased from **0.72 to 0.85**, showing an improved balance between Precision and Recall.

🔷 Key Performance Metrics

✅ **Confusion Matrix** – Showed correct vs. incorrect predictions.

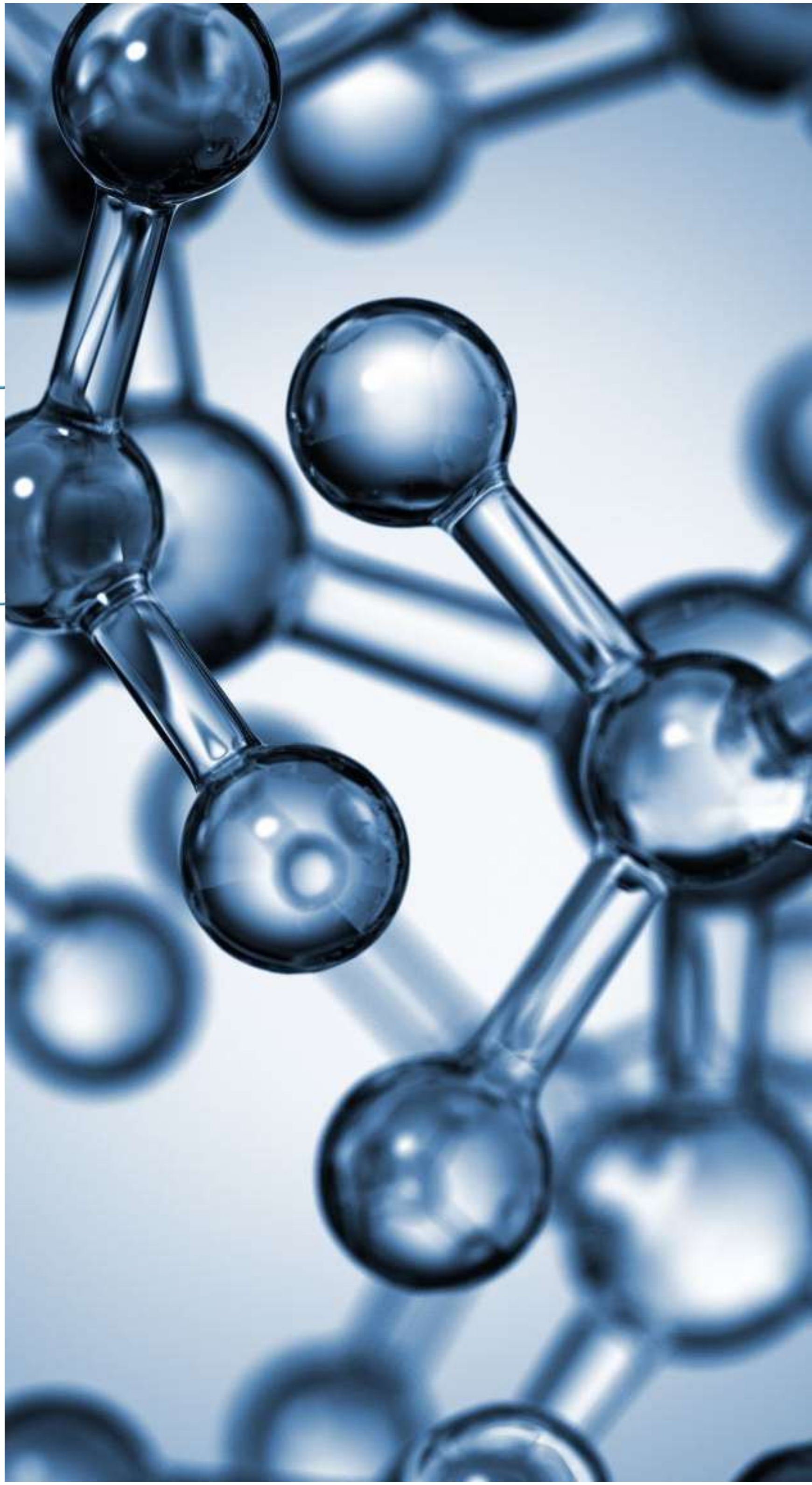
✅ **ROC Curve** – Confirmed the model's strong ability to distinguish between arrests and non-arrests.

✅ **Precision-Recall Curve** – Evaluated how well the model minimizes False Positives and False Negatives.

✅ **Bias Testing** – Ensured that the model **performed fairly across different demographic groups**.

✅ **Impact of Fine-Tuning** – After additional refinements, the model exhibited **better fairness balance** across race and gender.

Recommendations



Enhancing Law Enforcement Decision-Making

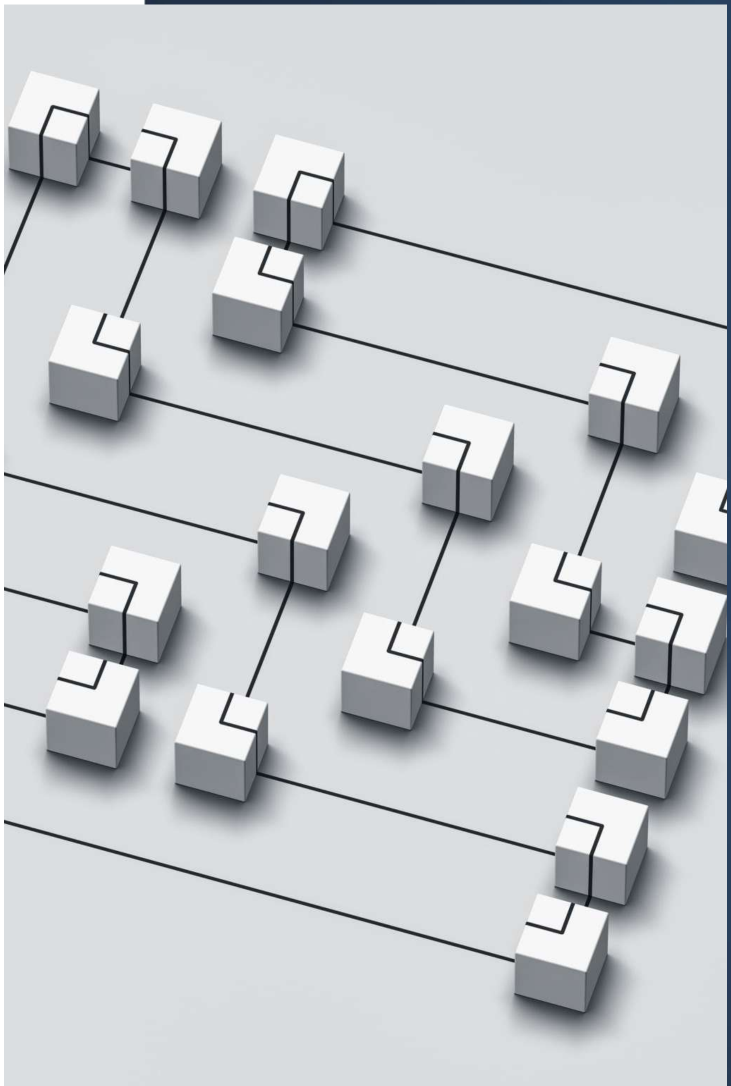
- ✅ **Use model insights to train officers** on improving stop justifications.
- ✅ **Deploy predictive analytics** to support real-time law enforcement decisions.
- ✅ **Ensure fairness audits** to prevent potential biases in arrest predictions.

Further Model Improvements

- **Compare Logistic Regression with Ensemble Models** – Evaluate Random Forest and Gradient Boosting for better performance.
- **Implement Deep Learning Techniques** – Explore Neural Networks to capture complex relationships in arrest patterns.
- **Continue Fine-Tuning Model Parameters** – Further optimize hyperparameters and decision thresholds.

Next Steps

- **Deploying the Model in Real-World Applications**



  - ✅ **Pilot the model in live policing environments** to assess practical effectiveness.
  - ✅ **Monitor predictions over time** to ensure continued accuracy and fairness.
  - ✅ **Work with law enforcement agencies** to integrate the model into decision-making frameworks.

🔷 Continuous Model Evaluation

✅ **Periodically reassess bias metrics** to maintain fairness in predictions.

✅ **Expand dataset coverage** by including factors like officer experience and community demographics.

✅ **Develop a monitoring system** to track **false positive and false negative rates** over time.

Conclusion

From this project I was able to;

- ✅ **Identify Key Factors Influencing Arrests .**

  The model successfully identified **Stop Resolution, Weapon Type,**and **Reported Hour** as the strongest predictors.

- ✅ **Improve Predictive Accuracy**

  Hyperparameter tuning and feature engineering **enhanced model performance**. Accuracy improved from **82% to 87%**, while the **F1-score increased from 0.72 to 0.85**.

- ✅ **Ensure Ethical & Fair Decision-Making**

  Bias testing revealed minor fairness variations across demographic groups. Further fairness audits are recommended to enhance **equitable law enforcement applications**.

◆ Policy & Ethical Considerations

For policy and ethical considerations;

- ✅ **Transparency & Explainability** – Ensure that law enforcement officers understand the model's decisions.
- ✅ **Regular Fairness Audits** – Conduct routine assessments to **detect and correct biases** in the model.
- ✅ **Stakeholder Collaboration** – Work closely with policymakers to ensure the model aligns with **legal and ethical standards**.

This project successfully built an **optimized and interpretable predictive model** to support data-driven law enforcement. By ensuring **ongoing fairness assessments, model monitoring, and ethical deployment**, agencies can **enhance transparency, accountability, and trust in predictive policing**.

Q&A?

Thank you !