ANALYZING PREDICTIVE FACTORS AND BIAS IN ARREST OUTCOMES DURING TERRY STOPS: A CLASSIFICATION APPROACH

•**Student Name:** MUTISYA FAITH MWENDE

Date: September 2024

Background

- **Terry v. Ohio (1968):** Established "reasonable suspicion" for Terry Stops, allowing temporary detainment of individuals based on suspicious behavior.
- **Objective:** Examine if race and gender biases influence arrest outcomes during Terry Stops.
- **Significance:** Understanding potential biases helps in promoting fair and transparent policing practices.

Problem Statement

- **Core Issue:** Investigate whether demographic factors such as race and gender affect the likelihood of an arrest during a Terry Stop.
- **Why It Matters:** Addressing potential biases ensures fairness in law enforcement and informs policy improvements.

Project Objectives

- **Data Exploration and Understanding**
  - Analyze key variables and address data quality issues.
- **Model Development**
  - Build and compare Logistic Regression and Decision Tree models.
- **Feature Importance Analysis**
  - Identify which factors are most influential in arrest outcomes.
- **Evaluation of Model Performance**
  - Assess accuracy, precision, recall, and F1-score.
- **Ethical Considerations and Recommendations**
  - Discuss ethical implications and suggest improvements for fairness.

Data Exploration

- **Dataset Overview:**
- **Key Columns:**
  - **Subject Age Group**
  - **Weapon Type**
  - **Officer ID, Officer Gender, Officer Race**
  - **Subject Perceived Race, Subject Perceived Gender**
  - **Arrest Flag**
- **Initial Data Checks:**
- **Missing Values:** Identified and addressed.
- **Statistics Summary:** Mean, standard deviation, min, max values

Data Cleaning

- **Missing Values:** Dropped null values to ensure data integrity.
- **Encoding Categorical Variables:**
- **Label Encoding:** Transformed categorical features into numerical values.
- **Preprocessing Steps:**
- **Numerical Features:** Standard scaling.
- **Categorical Features:** One-hot encoding.

Logistic Regression Model

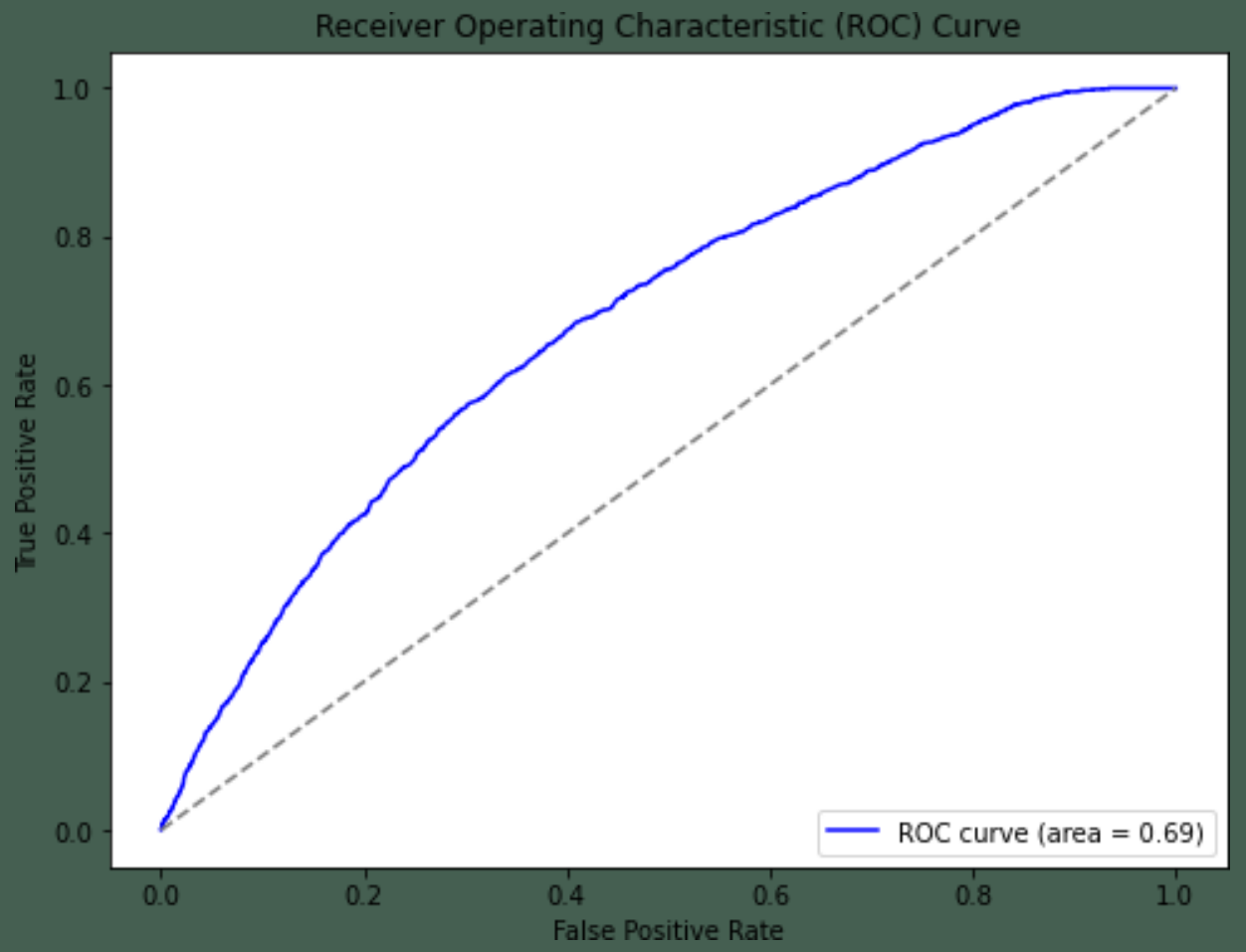
- **Model Details:**
- **Accuracy:** 89.57%
- **Precision, Recall, F1-Score:**
  - **For Negative Cases ("N"):**
    - **Precision:** 90%
    - **Recall:** 98%
    - **F1-Score:** 0.94
  - **For Positive Cases ("Y"):**
    - **Precision:** 24%
    - **Recall:** 5%
    - **F1-Score:** 0.08
- **Performance Insights:**
- Effective at predicting non-arrest cases.
- Challenges in predicting arrests indicate possible class imbalance.

Decision Tree Model

- **Model Details:**
  - **Overall Accuracy:** 88%
  - **Precision, Recall, F1-Score:**
    - **For Negative Cases ("N"):**
      - **Precision:** 91%
      - **Recall:** 96%
      - **F1-Score:** 0.93
    - **For Positive Cases ("Y"):**
      - **Precision:** 20%
      - **Recall:** 7%
      - **F1-Score:** 0.11
- **Feature Importance Analysis:**
  - Notable features affecting arrests: Demographics, weapon presence.
  - Inconsistencies in feature extraction noted.

ROC CURVE

- **ROC Curve Analysis:AUC:** 0.69, indicating moderate discriminative power.
- **Curve Insights:** Model is better than random guessing but shows room for improvement.


Receiver Operating Characteristic (ROC) Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.69)

Feature Importance

- **Decision Tree Insights:Most Influential Features:**
  - Demographic factors (race, gender).
  - Weapon presence.
- **Issues:**
  - Discrepancies in feature names.
  - Inaccurate importance scores due to data inconsistencies.

SUMMARY OF FINDINGS

- **Strengths:**
  1. High accuracy for predicting non-arrest cases.
  2. Good performance metrics for negative cases.
- **Weaknesses:**
  1. Poor performance in identifying arrest cases.
  2. Potential class imbalance impacting predictive accuracy.
- **Feature Impact:** Race, gender, and weapon presence are significant factors

Recommendations

- **Address Class Imbalance:**
- Use techniques like SMOTE or balanced class weights to improve prediction o
- **Ethical Considerations:**
- Implement fairness metrics and ensure transparency in predictive models.
- **Further Analysis:**
- Explore additional features such as time of day and geographical data.
- **Stakeholder Engagement:**
- Collaborate with law enforcement and community representatives to refine pra

CONCLUSION

- **Objective Achieved:** Analyzed predictive factors and potential biases in arrest outcomes during Terry Stops.
- **Key Insights:**
- Models show effectiveness in some areas but need improvement in others.
- Ethical and practical recommendations to enhance fairness and transparency.
- **Next Steps:** Implement recommendations, conduct further analysis, and engage stakeholders.

THE END