



0 stars

0 forks






1 watching

1 Branch

0 Tags

Activity

Private repository

	"Faith" Third commit	efdaa13 · 2 minutes ago	
	README.md	Third commit	2 minutes ago
	Terry_Stops_20240831.csv	data souce and notebook	yesterday
	index.ipynb	commit	5 minutes ago

# Phase-3-Project

## Analyzing Predictive Factors and Bias in Arrest Outcomes During Terry Stops: A Classification Approach

Student Name: Mutisya Faith Mwende  
Student Pace: Part-time

### Background

The Terry v. Ohio case established the concept of "reasonable suspicion," permitting police officers to briefly detain individuals based on suspicious behavior without probable cause. These stops, known as Terry Stops, are common in policing, particularly for suspected criminal activity. However, concerns about potential biases—particularly related to race and gender—have emerged. This project aims to use machine learning to predict arrest outcomes during Terry Stops and explore how demographic factors influence these outcomes.

### Problem Statement

The project investigates whether demographic factors, such as race and gender, influence the likelihood of arrest during a Terry Stop. Understanding these factors is crucial for assessing whether policing practices are equitable and free from bias. The goal is to develop a classification model to predict arrest outcomes based on various features and identify any potential biases.

### Objectives

- Data Exploration and Understanding:** Analyze key variables, including demographics, weapon presence, and time of day, and address any data quality issues.

2. **Model Development:** Build and evaluate classification models using logistic regression and decision trees to predict arrest outcomes.
3. **Feature Importance Analysis:** Identify influential features in predicting arrest outcomes, with a focus on demographic factors.
4. **Evaluation of Model Performance:** Use classification metrics to assess model accuracy, precision, recall, and F1-score, and compare performance between models.
5. **Ethical Considerations and Recommendations:** Discuss potential biases in the use of demographic data and provide recommendations for fairer policing practices.

## Data Exploration and Understanding

- **Dataset:** The dataset includes attributes such as subject age group, officer details, stop resolution, and arrest indicators.
- **Missing Values:** Identified and addressed through data cleaning processes.
- **Feature Analysis:** Key features include perceived race and gender of subjects, officer demographics, weapon type, and time of day.

## Data Cleaning and Preprocessing

- **Missing Values:** Dropped null values.
- **Categorical Encoding:** Converted categorical variables to numerical values using label encoding.
- **Feature Scaling and Transformation:** Applied standard scaling and one-hot encoding to numerical and categorical features respectively.

## Modeling

### 1. Logistic Regression:

- **Accuracy:** 89.57%
- **Performance:** High accuracy for negative cases, lower for positive cases, suggesting potential class imbalance.

### 2. Decision Tree:

- **Accuracy:** 88%
- **Performance:** Similar to logistic regression, effective for negative cases but less accurate for positive cases.
- **Feature Importance:** Provided insights into the influence of various features on arrest outcomes.

## Visualizations

- **Confusion Matrix Heatmap:** Shows strong performance for 'N' (negative) cases and weaker performance for 'Y' (positive) cases.
- **Precision, Recall, and F1-Score Bar Plot:** Highlights high precision and recall for negative cases and low scores for positive cases.
- **ROC Curve:** An AUC of 0.69 indicates moderate model performance, with room for improvement.

## Conclusion and Recommendations

**Conclusion:** The models performed well in identifying negative cases but struggled with positive cases. This suggests an imbalance in the dataset or challenges in predicting positive cases.

**Recommendations:**

1. **Address Class Imbalance:** Use techniques such as resampling or algorithms designed for imbalanced data.
2. **Ethical Considerations:** Incorporate fairness metrics and advocate for transparency in policing practices.
3. **Further Analysis:** Investigate the impact of each demographic variable and explore additional factors like time and location.
4. **Stakeholder Collaboration:** Engage with law enforcement and community representatives to discuss findings and develop actionable strategies.

Implementing these recommendations can enhance model effectiveness and contribute to more equitable policing practices.



## Releases

No releases published

[Create a new release](#)

## Packages

No packages published

[Publish your first package](#)

## Languages

● Jupyter Notebook 100.0%