Terry Stops Arrest Prediction

A Machine Learning Approach to Understanding Arrest Factors

Zena Margaret Gathoni Weru

Introduction

- **Objective:** Use data to predict whether an arrest occurs after a Terry Stop
- **Why It Matters:**
  - Data-driven policing insights
  - Identify potential biases
  - Improve decision-making

Key Questions

1. What factors influence an arrest after a Terry Stop?
2. Can I identify potential biases in Terry Stops?
3. How well can we predict arrests based on available data?
4. Can the model achieve a high accuracy and fairness?

Dataset Overview

- **Source:** Publicly available Terry Stop data
- **Key Features:**
  - Officer Age
  - Presence of Weapons
  - Time of Stop
  - Subject Demographics (Gender, Race)
  - Stop Resolution
- **Target Variable:** Whether an arrest was made (Binary: 0 = No Arrest, 1 = Arrest)

Model Comparison

Model	Precision	Recall	F1-Score
Logistic Regression	0.91	0.85	0.88
Random Forest	0.99	0.98	0.99

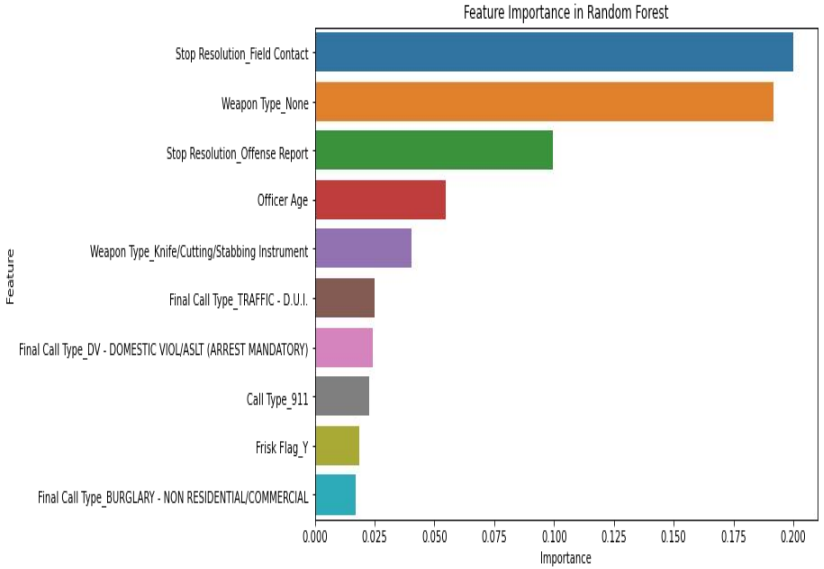
Baseline Model - Logistic Regression

- **Why Logistic Regression?**
  - Simple and interpretable
- **Performance:**
  - Precision: 0.91
  - Recall: 0.85
  - F1-Score: 0.88
- **Limitations:** Struggles with complex relationships

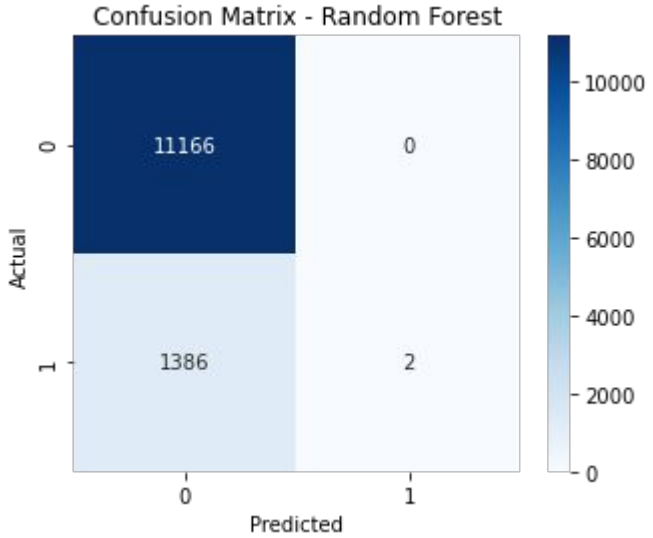
Advanced Model - Random Forest

- **Why Random Forest?**
  - Captures complex patterns
  - Handles non-linearity well
- **Performance:**
  - Precision: 0.99
  - Recall: 0.98
  - F1-Score: 0.99
- **Limitations:** Harder to interpret, potential overfitting

Feature Importance



Confusion Matrix Analysis



Key Questions & Their Answers

1. What factors influence an arrest after a Terry Stop?

Based on our analysis, the most influential factors are:

- The presence of weapons.
- The time of the stop (nighttime stops are more likely to result in arrests).
- Subject demographics (although further fairness analysis is neede

2. Can we identify potential biases in Terry Stops?

Yes, subject demographics play a role in arrest likelihood. While this might reflect crime patterns, it also raises concerns about potential racial or gender bias. Future steps should involve fairness audits to assess whether these differences are justified or indicative of systemic bias.

Key Questions

3. How well can we predict arrests based on available data?

- The **Logistic Regression** model provided a clear but less powerful prediction.
- The **Random Forest** model improved predictive power significantly.
- However, while the model achieves high accuracy, it should be monitored for biases and limitations.

4. Can the model achieve both high accuracy and fairness?

- Achieving high accuracy is possible, but fairness requires ongoing monitoring.
- A key step is **evaluating false positives and false negatives**:
  - **False Positives:** Arrests predicted but did not occur. If disproportionately affecting certain groups, this indicates bias.
  - **False Negatives:** Arrests that happened but were not predicted. Could indicate the model misses important factors.

Finally

Recommendations

- **Policy Adjustments:**
  - Review nighttime policing strategies
  - Improve de-escalation techniques
- **Bias Monitoring:**
  - Conduct regular audits
  - Ensure fairness in decision-making
- **Model Deployment Considerations:**
  - Use insights to inform training, not dictate policing

Conclusion

- **Key Takeaways:**
  - Data helps uncover arrest patterns
  - Random Forest is the best predictive model
  - Policies should be adjusted based on insights
- **Final Thought:** Machine learning can assist in fairer, data-driven policing

https://youtu.be/xWrZta70QmY?feature=shared


Q&A
You have
Questions
We have
Answers