Data-Driven Prediction of Reoffending Patterns in Charleston, South Carolina

COLLEGE OF CHARLESTON
SCHOOL OF ENGINEERING, COMPUTING, AND MATHEMATICS

Owayne Owens - owenso@g.cofc.edu

COLLEGE OF CHARLESTON
MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS

INTRODUCTION

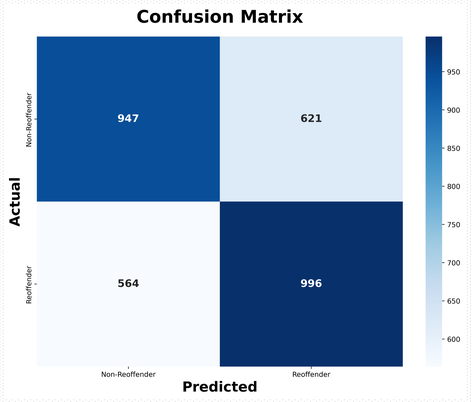
- Recidivism (reoffending) is a major concern for the criminal justice system, with implications for public safety and police resource allocation.
- Identifying individuals at higher risk of reoffending can support smarter interventions and rehabilitation efforts and create a safer society.
- This project applies machine learning to classify individuals as likely or unlikely to reoffend, using arrest record data from the **Charleston Police Department**.

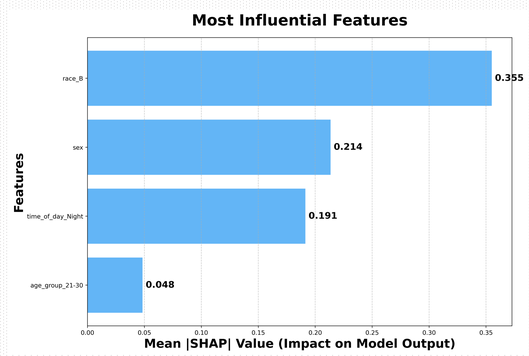
METHODS

- Dataset: 28,506 arrest records from the **Charleston Police Department**, including 7,592 labeled reoffenders.
- Highly correlated features (correlation > 0.75) were removed to reduce multicollinearity and improve interpretability.
- Class imbalance was addressed by under-sampling the non-reoffender class to match the number of reoffenders.
- Model: **Extreme Gradient Boosting (XGBoost)** with early stopping, max-depth tuning, and scale_pos_weight adjustment to mitigate class imbalance.
- Hyperparameters were tuned via cross-validation on an 80/20 train-test split (random seed = 42) to prevent overfitting.
- Evaluation prioritized **Precision, Recall, Accuracy, and Confusion Matrices** due to the high social cost of false positives and false negatives.
- **SHAP (SHapley Additive exPlanations)** was used to interpret model predictions and identify key predictors of recidivism.

RESULTS

- The objective was to predict the likelihood of reoffending, prioritizing recall to enhance the identification of potential reoffenders.
- Model performance:
  - **Recall**: 64% — correctly identified 64% of actual reoffenders.
  - **Precision**: 62% — 62% of predicted reoffenders were correctly classified.
  - **Accuracy**: 62% — overall classification accuracy.
- The confusion matrix showed the model identified most reoffenders but produced a considerable number of false positives.
- SHAP analysis identified race as the most influential predictor, followed by sex.
- Despite moderate predictive performance, bias and class imbalance remained key challenges.





Metric	Training	Testing
Recall	0.635	0.638
Precision	0.604	0.616
Accuracy	0.623	0.621

DISCUSSION & LIMITATIONS

- SHAP values provided interpretability by quantifying each feature's contribution to predictions; race emerged as the most influential predictor, raising concerns about bias and fairness.
- Initial models without resampling under-predicted reoffenders due to class imbalance, favoring the majority non-reoffender class.
- Under-sampling improved recall and balanced predictions, though false positives remained prevalent.
- XGBoost outperformed other models in handling tabular data and class imbalance but remained constrained by limited feature availability.
- The absence of key recidivism predictors such as prior convictions, substance use, employment status, and social support likely limited model accuracy and fairness.

CONCLUSIONS & NEXT STEPS

- Predicting recidivism remains challenging, particularly due to the absence of key social, economic, and behavioral features.
- The model achieved moderate performance but was sensitive to class imbalance and potential bias.
- SHAP analysis identified race as a strong predictor, highlighting fairness concerns in the model's decision-making.

For future work, I will:

- Try to enrich the dataset with features such as prior criminal history, employment status, family structure, and substance abuse history.
- Implement formal bias mitigation techniques (e.g., reweighting, fairness-aware modeling) to reduce disparate impact.
- Incorporate fairness metrics (e.g., demographic parity, equalized odds) to complement traditional performance evaluations and assess social implications.

REFERENCES

1. Charleston Police Department, Arrest Data - https://charleston-sc.maps.arcgis.com/apps/MapJournal/index.html?appid=1d87740b38ac4aa89795e87f7bc0faf6
2. Buikhuisen, W., & Hoekstra, H. A. (1974). FACTORS RELATED TO RECIDIVISM. *The British Journal of Criminology*, *14*(1), 63–69. http://www.jstor.org/stable/23636091
3. Glazebrook, J. S. (2009). Risky Business: Predicting Recidivism. *Psychiatry, Psychology and Law*, *17*(1), 88–120. https://doi.org/10.1080/13218710903040421