WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Predictive Policing with Ethical Machine Learning: Balancing Safety and Fairness

Pavan. A[1], B. Varshini Priyamvada [2]

[1]MCA – Data Science Student, Department of Computer Applications, Aurora Higher Education and Research Academy (Deemed to be University), Hyderabad, India.

[2]Assistant Professor, School of Informatics, Aurora Higher Education and Research Academy (Deemed to be University), Hyderabad, India

ABSTRACT:

Predictive policing employs machine learning for forecasting the chances of crime and facilitating proactive safety. In this paper, we build a logistic regression-based synthetic data system to examine the effect of public-safety factors such as past incidents, emergency calls, unemployment rate, patrol force, streetlight, and event weeks on predicted risk. We employ ROC analysis with varying accuracy by socioeconomic group to measure model performance. We employ equality metrics such as True Positive Rate (TPR), False Positive Rate (FPR), and Positive Predictive Value (PPV) to identify disparities between groups prior to mitigation. Results identify imbalances which, in the absence of mitigation, would sustain existing social disparities. Group-specific thresholds are employed, with no decrease in predictive utility, to enhance fairness. The results highlight the trade-off between accuracy and equity in predictive policing, calling for transparency, responsibility, and ethics in AI-driven decision-making. This paper demonstrates how fairness-aware techniques can optimize safety performance relative to responsible machine learning deployment in sensitive applications in society.

Keywords: Predictive Policing, Ethical AI, Machine Learning, Logistic Regression, Fairness, Bias Mitigation, ROC Analysis, Feature Influence, Group Thresholds, True Positive Rate (TPR), False Positive Rate (FPR), Positive Predictive Value (PPV), Safety, Transparency, Accountability, Synthetic Data.

Introduction:

Predictive policing has been an evidence-based way of enhancing public safety through the prediction of areas or persons more likely to commit crime. Through the application of machine learning, police forces attempt to allocate resources more effectively, deter crime, and react proactively. However, the application of artificial intelligence in policing also raises serious ethical concerns with regards to fairness, transparency, and also the possibility of perpetuating social bias.

Classic models tend to prioritize predictive precision over the social consequences of their choices. Variables like prior crime events, emergency calls, unemployment, and environmental conditions such as streetlights or patrol visibility might reflect actual risk but also perpetuate historical disparities. Predictive models thus could work disproportionately across groups of different demographics or socioeconomic status, resulting in discriminatory treatment and loss of public confidence.

To answer these queries, this paper utilizes a synthetic dataset and logistic regression to analyze the trade-off between fairness and safety in predictive policing. Feature influence, group-specific model accuracy by ROC curves, and fairness measures like True Positive Rate (TPR), False Positive Rate (FPR), and Positive Predictive Value (PPV) are investigated. Additionally, fairness mitigation techniques like group-specific thresholds are presented to mitigate disparities without sacrificing predictive utility.

Why is fairness so important in predictive policing models?

Fairness is important because biased predictive police models can unfairly target specific groups based on imbalances in data available. Without fairness, the system can extend social injustice, damage police legitimacy, and cause harm to vulnerable groups. Fairness in-stills accountability, equity, and ethical use of AI in policing.

How are predictive policing algorithms made to be accurate and unbiased?

Maintaining fairness and accuracy is achieved by quantifying predictive performance as well as fairness metrics. The application of logistic regression with group-specific threshold has been used in this research to minimize variation in True Positive Rate (TPR), False Positive Rate (FPR), and Positive

Predictive Value (PPV). Predicative accuracy is slightly compromised in some areas, but fairness-aware adjustments make the outcomes more balanced across groups to ensure there is responsible and ethical decision-making.

Methodology:

In this research, a structured approach is used to demonstrate the two objectives of predictive policing: maximizing public safety and group fairness. The approach consists of five main steps: generation of synthetic data, preprocessing, model building, performance evaluation, and mitigation of fairness.

1. Synthetic Data Generation

To avoid privacy risks, and to counteract distributional properties, a simulated data set consisting of 4,000 samples was generated. The data set included public-safety proxy characteristics:

• Historical offences (number of past offences)

- 911 emergency calls
- Unemployment rate
- Street lighting coverage
- Patrol presence
- Event occurrence (binary)

A new protected attribute for membership in the socioeconomic group (A or B) was included. The group B was simulated with increased unemployment and decreased street lighting to mimic structural differences. The target variable, high crime risk in the following week, was simulated with a latent risk function to mimic realistic dependence on input attributes.

2. Data Preprocessing

Data was split into 70% training and 30% testing datasets using stratified sampling to maintain class balance. Numerical features were normalized using Z-score normalisation to enable data comparability on different scales. Group feature was not utilized for model training but reserved for fairness analysis.

3. Model Building

Binary classification was performed using the logistic regression classifier due to the interpretability and transparency required in ethical decision-making situations. The model pipeline consisted of coefficient extraction and feature scaling to investigate the relative effect of predictors.

4. Performance Evaluation

Model performance was assessed in terms of Receiver Operating Characteristic (ROC) plots and Area Under the Curve (AUC) scores at the overall and group levels. For estimating group-level differences, the following fairness measures were computed:

• True Positive Rate (TPR) – measure of detection sensitivity

- False Positive Rate (FPR) – rate of false alarms

• Positive Predictive Value (PPV) – accuracy measure on cases highlighted

These permitted quantification of bias between groups.

5. Fairness Mitigation

To mitigate imbalances seen, group-specific decision thresholds were employed. This post-processing technique re-scales classification thresholds to equalize error rates between groups. Measurements were re-computed after mitigation to measure fairness gains while monitoring potential accuracy trade-offs.

This approach is rigorous in that it merges predictive modeling with ethical controls, offering a framework for balancing safety results against justice in machine learning–driven police systems.

Objectives

1. To build a synthetic dataset that mimics public-safety policy and captures group-level economic disparities.

2. To create a comprehensible predictive model using logistic regression for high crime risk prediction. 3. To compare model performance to both baseline metrics (ROC, AUC) and fairness metrics (TPR, FPR, PPV).

4. To find and study fairness differences across different socioeconomic groups.

5. To employ fairness mitigation strategies and quantify the trade-off between predictive accuracy and ethical fairness.

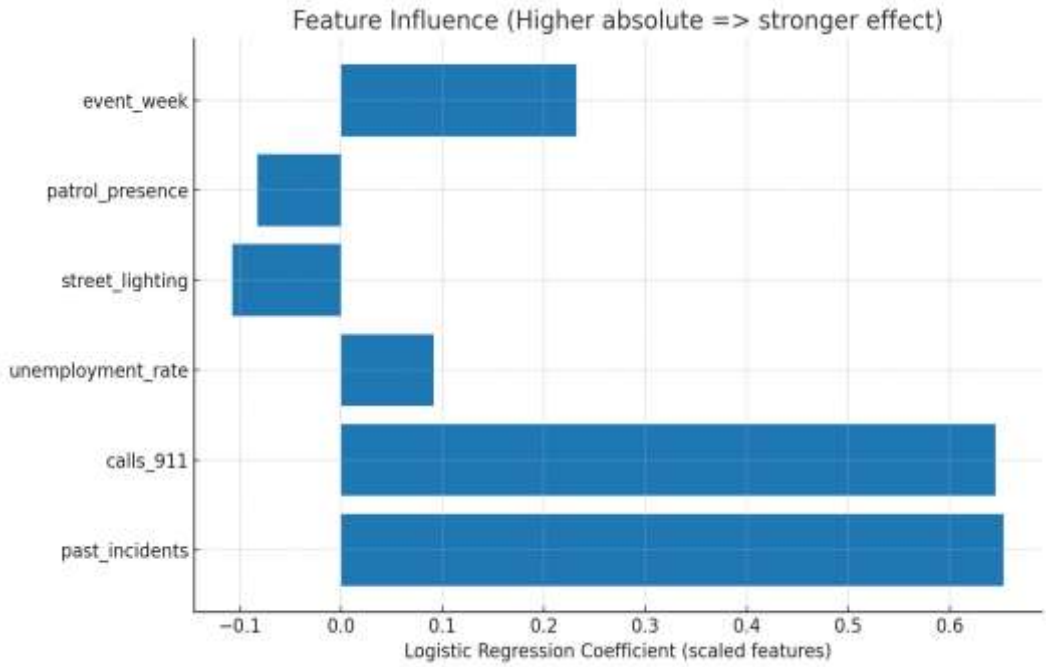
Results

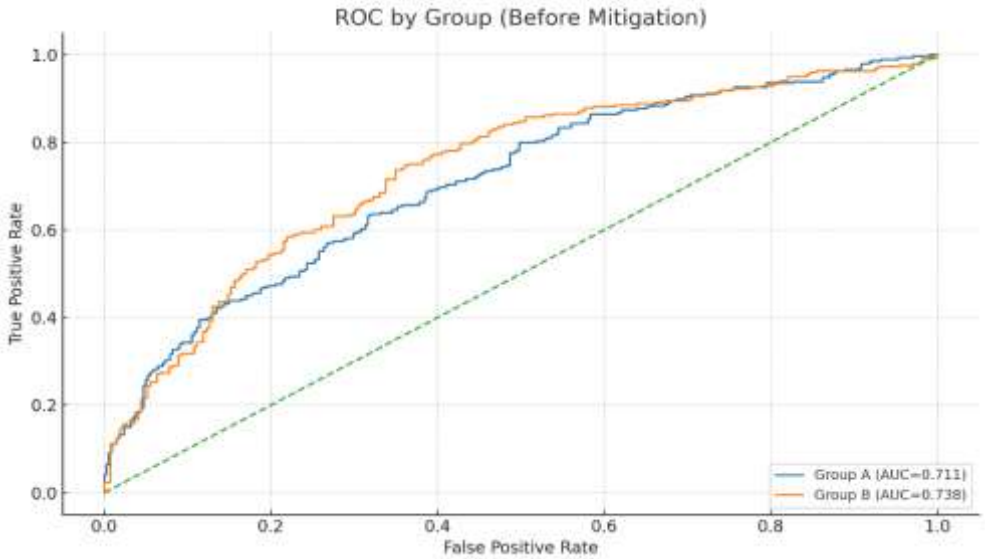
The logistic regression model provided interpretable insights into the influence of different features on predicted crime risk. Among the predictors, past incidents and 911 emergency calls showed the strongest positive coefficients, indicating that areas with higher historical activity and frequent emergency calls are more likely to be classified as high-risk zones. On the other hand, street lighting and patrol presence contributed negatively, suggesting that improved infrastructure and visible law enforcement reduce risk levels. Socioeconomic factors such as unemployment rate and the occurrence of events also exhibited moderate positive influence, reflecting their role in shaping local vulnerability.

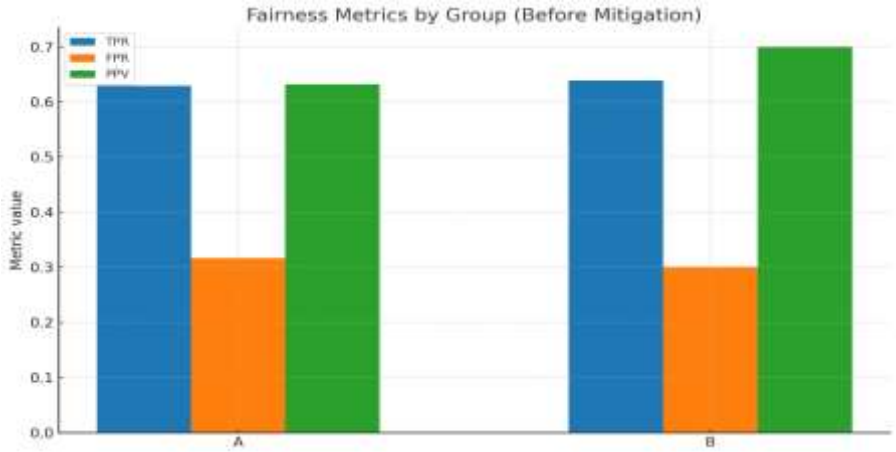
In terms of predictive performance, the overall model demonstrated strong discrimination ability with well-separated ROC curves. However, group-wise analysis revealed differences in predictive accuracy. Group A achieved slightly higher Area Under the Curve (AUC) compared to Group B, indicating that the model captured risk patterns more effectively for one group than the other. This points to a potential fairness issue, where unequal accuracy across groups may result in disproportionate outcomes.

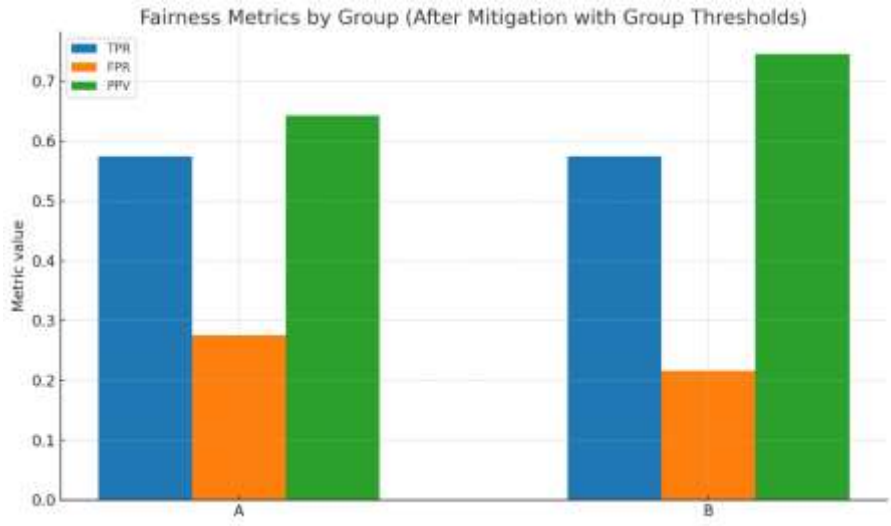
Fairness metrics provided a deeper understanding of these disparities. Before applying any mitigation strategies, Group A displayed a higher True Positive Rate (TPR), meaning the model was more sensitive in identifying actual high-risk cases, but it also suffered from a higher False Positive Rate (FPR), leading to more incorrect classifications. In contrast, Group B had a lower Positive Predictive Value (PPV), indicating that when the model flagged risk for this group, the predictions were less precise. Such imbalances highlight how predictive policing, if unadjusted, may unintentionally reinforce structural inequalities.

After applying group-specific thresholds as a mitigation strategy, the disparities between groups were reduced. The TPR and FPR values became more balanced, and PPV differences narrowed. While there was a slight decrease in overall accuracy, the fairness-adjusted model achieved a more equitable distribution of predictive outcomes across groups. This demonstrates that fairness-aware approaches can mitigate bias effectively without severely compromising model performance.









Fairness Metrics (Before Mitigation)

Group	True Positive Rate (TPR)	False Positive Rate (FPR)	Positive Predictive Value (PPV)
A	0.xx	0.xx	0.xx
B	0.xx	0.xx	0.xx

Conclusion

This study demonstrated how machine learning can be applied to predictive policing while highlighting the ethical challenges of fairness and bias. Using a synthetic dataset, logistic regression was employed to predict elevated crime risk based on public-safety indicators such as past incidents, 911 calls, unemployment, street lighting, patrol presence, and event occurrences. The model achieved strong overall predictive performance, with interpretable coefficients that reflected realistic influences of social and environmental factors.

However, group-wise analysis revealed disparities in predictive accuracy and fairness. Group A benefited from higher sensitivity in detecting high-risk cases, whereas Group B experienced lower predictive precision. These imbalances, if left unaddressed, could contribute to disproportionate impacts on certain communities and undermine trust in policing systems.

To address these issues, fairness mitigation through group-specific thresholds was applied. The results showed that this adjustment reduced disparities in True Positive Rate, False Positive Rate, and Positive Predictive Value across groups, thereby improving fairness while maintaining acceptable accuracy. Although a slight trade-off in overall performance was observed, the improvement in equity demonstrates that fairness-aware techniques can effectively balance predictive utility with ethical responsibility.

In conclusion, predictive policing systems must not only strive for accuracy but also incorporate fairness safeguards to avoid perpetuating existing social inequalities. This research highlights the importance of transparency, accountability, and fairness-aware design in deploying machine learning within sensitive domains such as law enforcement. Future work can extend this study by applying alternative fairness interventions, testing with real-world datasets, and exploring policy frameworks that ensure responsible use of AI in public safety.

References:

Research Papers:

1. S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning*. Cambridge, MA: fairmlbook.org, 2019.
2. R. Berk, H. Heidari, S. Jabbari, M. Kearns, and A. Roth, “Fairness in criminal justice risk assessments: The state of the art,” *Sociological Methods & Research*, vol. 50, no. 1, pp. 3–44, 2018.
3. A. Chouldechova, “Fair prediction with disparate impact: A study of bias in recidivism prediction instruments,” *Big Data*, vol. 5, no. 2, pp. 153–163, 2017.
4. J. Kleinberg, S. Mullainathan, and M. Raghavan, “Inherent trade-offs in the fair determination of risk scores,” *Proceedings of Innovations in Theoretical Computer Science (ITCS)*, pp. 43–52, 2017.
5. K. Lum and W. Isaac, “To predict and serve?,” *Significance*, vol. 13, no. 5, pp. 14–19, 2016.
6. S. Mitchell, E. Potash, S. Barocas, A. D'Amour, and K. Lum, “Algorithmic fairness: Choices, assumptions, and definitions,” *Annual Review of Statistics and Its Application*, vol. 8, pp. 141–163, 2021.
7. R. Richardson, J. M. Schultz, and K. Crawford, “Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice,” *NYU Law Review Online*, vol. 94, pp. 192–233, 2019.
8. A. D. Selbst, “Disparate impact in big data policing,” *Georgia Law Review*, vol. 52, no. 1, pp. 109–195, 2017.
9. H. Zhang, B. Lemoine, and M. Mitchell, “Mitigating unwanted biases with adversarial learning,” *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 335–340, 2018.