A Decision Tree Analysis of Bias in Predictive Policing: A Cyber Law Perspective

Liu Yang[1,*], Matee Pigultong[2]

[1]Vocational Education Division, Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Thailand

[2]Educational Technology and Communications Division, Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Thailand

ABSTRACT

As law enforcement agencies increasingly adopt data-driven technologies, predictive policing systems present a significant challenge to constitutional principles of equal protection and due process. This paper provides an interdisciplinary analysis of this challenge from a cyber law perspective, using a machine learning model as a technical case study. The primary objective was to build an interpretable predictive policing model and audit it for both performance and encoded demographic bias, thereby creating a concrete foundation for a legal and ethical critique. A Decision Tree classifier was trained on a publicly available Crime & Safety dataset to predict crime types based on a combination of temporal, geographical, and victim demographic data. The methodology involved standard data preprocessing, feature engineering, and the use of a "white-box" model specifically chosen for its high degree of interpretability. The model's performance was evaluated using standard metrics, including accuracy, precision, and recall, while bias was assessed through a detailed analysis of feature importances and a direct inspection of the tree's decision-making logic. The results demonstrated a dual failure of the model. First, it was functionally ineffective, achieving an overall accuracy of only 10%, rendering it useless for practical application. Second, and more critically, the feature importance analysis revealed that the model was systematically biased, relying heavily on protected characteristics such as victim race and gender to make its classifications. The visualization of the Decision Tree provided direct, irrefutable evidence that these demographic factors were used to create explicit decision-making rules within the algorithm. This study concludes that the deployment of such an unaudited model would be both negligent, due to its inaccuracy, and unconstitutional, due to its discriminatory logic. The findings illustrate the profound legal risks municipalities face and underscore the absolute necessity of mandatory, independent audits and public transparency reports before any predictive policing system is deployed. The interpretability of the model proved to be a powerful tool for exposing bias, highlighting the importance of Explainable AI (XAI) in the legal oversight of algorithmic governance.

Keywords Algorithmic Bias, Cyber Law, Decision Tree, Disparate Impact, Predictive Policing

Submitted 7 July 2025
Accepted 24 July 2025
Published 1 September 2025

*Corresponding author
Liu Yang,
liu.yang01@gmail.com

Additional Information and Declarations can be found on page 226

DOI: 10.63913/jcl.v1i3.12

Introduction

Central to predictive policing is the application of algorithmic models that leverage vast amounts of data collected from various sources. These models analyze historical crime data, socioeconomic indicators, and even real-time reports from social media platforms to establish patterns and predict where and when crimes are likely to occur [1][2]. Notably, predictive policing employs both geospatial and individual-based approaches, allowing law enforcement to forecast crime in specific locations or target individuals deemed at risk [3]. This dual categorization reflects the diversity of predictive policing tools and the wide-

ranging implications they may have on community interactions and civil liberties [4][5].

Despite its potential benefits, predictive policing is not without controversy. The use of algorithms raises concerns about transparency, accountability, and the potential perpetuation of existing biases within the criminal justice system. Algorithms often function as 'black boxes,' meaning the processes and rationale behind their predictions are not readily apparent, which can lead to criticisms regarding their fairness and ethical implications [6][4]. Research indicates that when predictive policing systems are trained on biased data, they risk exacerbating disparities in law enforcement outcomes, particularly concerning marginalized communities [4][7]. This prominence of bias highlights the necessity for ongoing scrutiny and reform to ensure that predictive policing does not inadvertently foster discriminatory practices.

The implementation of predictive policing is particularly evident in large metropolitan areas, where the analysis of big data can significantly enhance policing efficiency. Implementations in cities like Los Angeles have reportedly resulted in improved crime prevention rates compared to traditional approaches [1][2]. Furthermore, studies indicate that departments focusing on real-time data processing exhibit significant improvements, such as a reported 42% enhancement in crime prevention rates and a reduction in response times [1]. This efficacy suggests that integrating algorithmic tools into everyday policing operations facilitates a paradigm shift toward a data-centric law enforcement model that prioritizes intelligence over intuition [2][8].

Moving beyond efficiency, predictive policing also aims to redefine police-community relationships by fostering greater cooperation and trust. The systematic use of data-driven strategies can aid in the timely dispatch of officers to high-risk areas, potentially reducing crime and increasing public perception of safety. However, the culture of control intrinsic to predictive policing initiatives requires careful management to mitigate public anxiety over increased surveillance and the potential erosion of civil liberties [9][7]. As communities grapple with the complexities of predictive policing, the dual edges of protection and potential overreach become essential conversational threads in discussions about the future of policing.

As this data-driven approach continues to permeate policing structures, there are broader societal implications to consider. The advent of predictive policing not only marks a shift in operational methodology but also signifies an evolving synergy between technology and social governance. The integration of predictive analytics within law enforcement thus embodies broader societal values, highlighting the delicate balance between maintaining public safety and ensuring the protection of civil rights [5][10]. This dynamism suggests that while predictive policing provides tools that can be beneficial, they must be wielded with care and rigorous oversight to uphold ethical standards within the realm of law enforcement.

The trajectory of predictive policing indicates a future where data analytics will become increasingly indispensable to crime prevention strategies. Yet, as jurisdictions embrace these innovations, they must weigh the benefits against the ethical and societal challenges presented by algorithmic governance. Collaboration between law enforcement agencies, policy-makers, and community stakeholders will be pivotal to promote transparent dialogues surrounding the use of predictive policing technologies [11]. Ultimately,

engaging in these essential conversations will ensure that law enforcement practices do not just adapt to new technologies but also honor the principles of justice and equity central to democratic societies.

The emergence of AI technologies necessitates a reevaluation of existing legal frameworks to address concerns surrounding discrimination and biased outcomes in predictive policing. Algorithms can inadvertently reflect societal biases, as historical data often encapsulates deep-rooted prejudices against specific racial or socioeconomic groups. Therefore, state actors are not only obligated to comply with the Equal Protection Clause but also to actively mitigate the risks that AI presents to equity in law enforcement. A systematic approach to understanding and dismantling these risks is crucial, as evidenced by ongoing legal and academic discourse surrounding the intersection of AI and human rights protections.

Emerging AI governance frameworks contribute significantly to establishing standards aimed at regulating AI in law enforcement. The introduction of comprehensive legislation, such as the European Union's AI Act, exemplifies the global movement toward crafting a robust regulatory approach that places human rights at the forefront of AI deployment. The AI Act positions itself as a critical system for assessing and categorizing AI applications based on their risk levels, thereby addressing concerns that align with the Equal Protection Clause by enforcing compliance protocols that governments must follow to protect their citizens. Moreover, these frameworks underscore the need for accountability mechanisms that enable oversight over algorithmic decision-making processes within policing operations, thus aligning with the constitutional mandate to uphold equal protection under the law.

In parallel, important conversations surrounding the ethical design of AI technologies are emerging, advocating for the integration of fundamental rights directly into the development of AI systems. Inclusive frameworks that recognize vulnerabilities and define the applicability of human rights principles in AI applications are paramount in addressing biases that may disenfranchise certain groups. The necessity of weaving human rights considerations into AI governance cannot be overstated, as it reinforces the collaborative responsibility of state actors to uphold dignity and fairness for all individuals irrespective of their backgrounds.

The multifaceted challenges posed by AI in the context of the state’s legal obligations extend into discussions about police conduct, public oversight, and the relationship between law enforcement agencies and the communities they serve. Legal consciousness among community members regarding their rights—especially in the face of technology-driven enforcement—is crucial. Awareness, education, and advocacy play central roles in equipping individuals to assert their rights under the Fourteenth Amendment, emphasizing the vital connection between public engagement and regulatory accountability.

This paper pursues a dual objective, operating at the intersection of data science and legal scholarship. The primary goal is to conduct a technical evaluation of a Decision Tree machine learning model to determine its potential for encoding and perpetuating demographic bias. Using the public Crime & Safety Dataset as a basis for this analysis, the study will construct a predictive policing model and perform a rigorous audit of its performance, feature importances, and internal decision-making logic. This technical investigation is designed to provide empirical, interpretable evidence of how a seemingly neutral algorithm

can learn and operationalize biases present in historical data.

Flowing from this technical analysis, the second objective is to analyze the profound implications of these findings for cyber law and constitutional rights. The paper will connect the quantitative results—specifically, the model's reliance on protected characteristics like race and gender—to established legal doctrines such as the Equal Protection Clause and the concept of disparate impact. By grounding the legal discussion in a concrete technical example, this study aims to move beyond theoretical debate to a practical demonstration of the legal risks and constitutional harms posed by unaudited predictive policing systems. The paper is structured accordingly: it will first review the relevant technical and legal literature, then detail the methodology for the model's construction and evaluation, present the results of the technical audit, and conclude with a discussion of their legal and policy ramifications.

Literature Review

Technical Foundations of Predictive Policing Models

The use of machine learning in predictive policing has become increasingly prominent, with Decision Trees emerging as key tools for crime prediction due to their interpretability and effectiveness. Decision Trees are algorithms that employ a flowchart-like structure to make decisions based on input features, which can include variables such as geographical information, socio-economic data, and historical crime records. Several studies have demonstrated the value of Decision Trees alongside other classifiers in predictive policing, emphasizing the importance of interpretable models in ensuring accountability and transparency in law enforcement actions [12][13].

Decision Trees are particularly advantageous because they allow for easy visualization of the decision-making process. Each node in the tree represents a feature, and the branches depict outcomes based on specific conditions related to that feature. Studies reveal that the interpretability of Decision Trees not only aids in justifying law enforcement decisions but also facilitates communication with policymakers and the public, which can enhance trust and support for predictive policing technologies [12]. This potential for interpretability addresses a criticism of machine learning models: their "black box" nature, which often makes it difficult for stakeholders to understand how decisions are derived [13].

In addition to Decision Trees, other machine learning classifiers such as Random Forests and Support Vector Machines (SVM) have been employed in crime prediction, showcasing varying degrees of predictive accuracy and interpretability. Random Forests enhance the robustness of predictions by aggregating the outputs of multiple Decision Trees, reducing the likelihood of overfitting and improving generalization on unseen data [14]. However, while more complex models like Random Forests may improve prediction accuracy, they often sacrifice some interpretability. Hence, the balance between model accuracy and transparency remains a critical factor in the adoption of these technologies by law enforcement agencies [13].

A specific challenge in implementing machine learning models involves preprocessing categorical data, which can include variables such as city, state, and victim race. One widely used approach for handling categorical data in machine learning is one-hot encoding. This method transforms categorical

variables into a binary array, where each category is represented by a distinct binary feature. For instance, if a categorical variable denotes the cities "New York," "Los Angeles," and "Chicago," one-hot encoding would result in three separate binary variables that indicate the presence or absence of each city for a given data point [15]. This technique is effective because it allows machine learning models to interpret categorical data numerically while retaining the nominal nature of the variable.

The process of one-hot encoding is instrumental in ensuring that machine learning models can utilize categorical data without imposing any unintended ordinality, which could lead to misinterpretations of the relationships among different categories [15]. However, an important limitation of this method is that it can lead to high dimensionality, particularly with features that have many unique categories (high-cardinality variables). This increased dimensionality can complicate model training and lead to inefficiencies; hence, alternative encoding strategies are also being explored [15].

The core formula governing Decision Trees, particularly in selecting the features for data splitting, is Information Gain. The concept of Information Gain quantifies the reduction in uncertainty about a target variable based on the information provided by a feature. It is calculated using the following formula:

$$Gain(S, A) = Entropy(S) - \sum v \in Values(A) \frac{|Sv|}{|S|} \cdot Entropy(S_v)$$

In this equation, (Gain(S, A)) represents the Information Gain associated with feature (A) on dataset (S). Here, (Entropy(S)) reflects the uncertainty in the original dataset (S), while the summation calculates the expected entropy of the subsets (S_v) created by splitting on feature (A). The lower the resulting entropy, the greater the Information Gain, indicating that feature (A) is a valuable predictor of the outcome.

Thus, Information Gain serves as a primary metric for guiding the partitioning of data within a Decision Tree framework. By continuously selecting features that provide the highest Information Gain, Decision Trees can efficiently and accurately derive predictions regarding crime occurrences based on historical patterns and variables.

Legal and Regulatory Landscape of Algorithmic Justice

The integration of algorithmic systems into public policy, particularly within law enforcement, raises vital legal and ethical considerations, most notably regarding issues of "disparate impact." Disparate impact refers to situations where a seemingly neutral policy disproportionately affects a particular group, even if there is no intention to discriminate. Discussions surrounding algorithmically driven predictive policing systems often highlight the risks of such technologies unintentionally perpetuating existing inequalities. Recent research indicates that legal systems are fundamentally designed to scrutinize both direct discrimination and its indirect implications, emphasizing the need for a comprehensive understanding of how algorithmic decision-making systems can lead to detrimental outcomes for specific demographic groups without overt bias [16][17].

Moreover, the use of AI and machine learning technologies in predictive policing implicates various data protection laws. The processing of sensitive demographic information—such as race, gender, and socioeconomic status—

poses complex challenges. Laws such as the General Data Protection Regulation (GDPR) in Europe, along with various national equivalents, set stringent requirements for how personal data is collected, processed, and stored [18]. Specific provisions, including the need for informed consent, data minimization, and purpose limitation, dictate how law enforcement agencies can utilize collected data, particularly when it encompasses sensitive characteristics. The enforcement of these laws requires organizations to ensure proper data governance and compliance with regulatory standards to avoid legal repercussions stemming from misuse or inaccurate processing of personal information.

Civil rights statutes play a pivotal role in challenging biased algorithmic systems employed by government entities. Notably, Title VI of the Civil Rights Act of 1964 prohibits federal programs from discriminating based on race, color, or national origin. These provisions create a legal foundation for individuals to contest algorithmic decision-making practices that yield discriminatory results—even if unintentional [17]. Moreover, recent proposals aimed at reforming existing anti-discrimination laws seek to adapt these legal frameworks to address the challenges posed by digital discrimination, particularly concerning socio-economic status [17][19]. Such reforms aim to ensure that vulnerable populations are protected from the adverse effects of algorithmically derived decisions.

Furthermore, the discussion surrounding AI systems in law enforcement extends to how civil rights statutes can be utilized to advocate for transparency and accountability. As algorithmic systems are increasingly employed, it is crucial that entities using these systems adhere to principles that allow individuals affected by algorithmic decisions the ability to understand, contest, and seek redress. This is essential in safeguarding civil liberties and mitigating the risks associated with opaque decision-making processes characteristic of many AI systems [20].

Elements of algorithmic accountability have garnered increasing attention as policymakers and legal scholars explore how to ensure that systems are not only effective in their objectives but also align with societal values and norms. Engaging in continual assessments and audits of these systems addresses concerns about biases, ensuring that they do not violate civil rights provisions while fostering greater trust in their utilization by law enforcement [20][21]. This emphasis on accountability highlights the necessity of developing regulatory frameworks that provide clear guidelines on the governance of algorithmic systems within public agencies.

Ultimately, the legal and regulatory landscape surrounding algorithmic justice necessitates a thoroughly considered approach that balances technological innovation with stringent protections for civil rights. Legal principles like disparate impact underscore the need for vigilance in deploying algorithmic systems to safeguard against unintended discrimination. The application of existing civil rights statutes, coupled with evolving data protection laws, provides avenues for challenging biases in these systems. As informed public discourse continues, the call for comprehensive regulations and accountability mechanisms becomes increasingly pressing, underscoring the importance of fostering equitable and just technological practices in law enforcement.

Method

This study employed a quantitative, computational approach to construct and critically evaluate a predictive policing model. The core of the methodology was to simulate the development of a data-driven system and then audit it for evidence of demographic bias, providing a technical foundation for the subsequent legal analysis. Using a publicly available Crime & Safety Dataset containing 1,000 records, the methodology was systematically divided into three distinct stages: comprehensive data preprocessing and feature engineering, in-depth exploratory data analysis to identify baseline disparities, and finally, the construction, training, and rigorous evaluation of a Decision Tree classification model.

Dataset and Preprocessing

The initial and most critical phase of the methodology involved the meticulous preparation of the dataset for analysis. The raw data, loaded into the analytical environment using the pandas library, represented a collection of simulated crime reports. Each record contained multiple fields, including the crime_type (the target variable for prediction), geographical information such as city and state, temporal data in date and time columns, and sensitive demographic details like victim_gender and victim_race. Recognizing that raw data is seldom suitable for direct input into machine learning algorithms, a series of preprocessing steps were executed.

A key step was feature engineering, a process of creating new, more informative features from the existing data. The original date and time columns were programmatically merged into a single, structured datetime object. This transformation was crucial as it allowed for the extraction of more granular temporal features that could plausibly correlate with criminal activity patterns. Four new features were engineered from this object: year, month, day_of_week, and hour. The rationale for this was to empower the model to learn potentially complex relationships, such as whether certain crimes are more prevalent during specific times of day or on weekends. Following this, a data cleaning step was performed to reduce noise and model complexity. Columns deemed redundant or irrelevant for the predictive task—such as the unique record id, the original date/time fields (now superseded by the engineered features), and the free-text location_description field—were permanently removed from the dataset. This left a refined set of features, including the categorical variables like city, state, victim_gender, and victim_race, which were specifically flagged for the necessary numerical transformation in the subsequent modeling stage.

Exploratory Data Analysis (EDA)

Before the construction of the predictive model, a thorough exploratory data analysis was conducted. The primary purpose of this stage was twofold: first, to gain preliminary insights into the dataset's underlying statistical properties, and second, to proactively identify and document potential sources of inherent bias that could be learned and amplified by the model. A central objective was to establish a clear, empirical baseline of the relationship between the target variable, crime_type, and the sensitive demographic attributes recorded in the data. To achieve this, a targeted visualization was generated using the seaborn and matplotlib libraries. A countplot was chosen to illustrate the absolute frequency distribution of different crime types across the various victim_race categories. This graphical analysis served as a foundational diagnostic tool, providing a transparent view of any significant disparities within the reported

data itself, before any predictive modeling took place. For instance, observing that a particular demographic group is disproportionately represented as victims for a specific crime type in the raw data is a critical finding. This initial analysis informs the subsequent interpretation of the model's results; if the model later relies heavily on race to predict that crime, the EDA provides evidence that this is likely a result of learning from skewed initial data rather than discovering a causally valid pattern.

Modeling and Evaluation

The predictive component of this study was a Decision Tree classifier. This specific algorithm was deliberately chosen over more complex, "black-box" models (such as neural networks or ensemble methods) due to its high degree of interpretability. For a study centered on auditing algorithmic bias for legal and ethical review, the ability to inspect the model's internal logic is not merely beneficial but essential. The entire modeling workflow was encapsulated within a scikit-learn Pipeline, a best practice that ensures procedural integrity by sequencing operations and preventing common errors like data leakage from the test set into the training process.

The dataset was first partitioned into a training set, comprising 70% of the records, and a testing set with the remaining 30%. A stratified sampling method was employed during this split to ensure that the proportional representation of each crime type was consistent across both the training and testing partitions. This is particularly important for datasets with imbalanced classes, as it prevents the possibility of a rare crime type being excluded from the test set entirely. Within the pipeline, a ColumnTransformer was used to apply one-hot encoding to the categorical features (city, state, victim_gender, victim_race). This process converts each category into a new binary column, transforming the non-numerical data into a format that the Decision Tree algorithm can process. The classifier was then trained on the preprocessed training data, with its max_depth hyperparameter constrained to 5 to prevent the model from becoming overly complex and memorizing the training data (overfitting).

Finally, the model's performance was rigorously evaluated on the unseen test data using a suite of metrics designed to provide a holistic view of its effectiveness and fairness. A detailed classification_report was generated to assess the precision (the accuracy of positive predictions), recall (the ability of the model to find all relevant instances), and F1-score (the harmonic mean of precision and recall) on a per-class basis. This granular analysis is critical for detecting bias, as a model can have high overall accuracy while still performing very poorly for specific subgroups. Additionally, a confusion matrix was created to visually represent the model's specific successes and failures in distinguishing between the different crime types.

Result and Discussion

This section presents the empirical results derived from the Decision Tree model, followed by a detailed technical and legal analysis of the model's performance and underlying logic. The findings indicate that the model is not only highly inaccurate but also systematically reliant on sensitive demographic data, which raises significant legal and ethical considerations. This dual failure—a complete lack of predictive utility that is compounded by the active encoding of societal biases—provides a compelling case study of the uniquely potent risks associated with deploying unaudited algorithmic systems in the public sector.

Performance of the Predictive Model

The quantitative evaluation of the model demonstrated a profound and unequivocal lack of predictive power. The overall accuracy on the unseen test set was a mere 10%, which indicates that the model incorrectly classified the crime type in 90% of instances. In a classification problem with ten distinct categories, this level of performance is statistically indistinguishable from random chance. Consequently, this result renders the model functionally inadequate for any practical law enforcement application, as its predictions offer no informational value beyond what could be achieved by a stochastic process.

A more granular analysis, detailed in the classification report, underscores this comprehensive failure. For several major crime categories—including Assault, Homicide, and Theft—the model's precision and recall were both 0.00. This signifies a total inability to identify even a single instance of these crimes, thereby creating significant predictive deficiencies. For most other categories, such as Arson, Drug Offense, and Robbery, the F1-scores were exceptionally low (ranging from 0.04 to 0.08), demonstrating a near-total inability to balance the competing demands of precision and recall. In a real-world context, this translates to a system that would simultaneously fail to identify the vast majority of crimes while also misclassifying the few it does predict.

The only crime type for which the model showed any meaningful, albeit weak, predictive capability was Burglary, which achieved a high recall of 0.61 but a dismal precision of only 0.15. This specific failure mode is particularly illustrative, as it shows that the model adopted a simplistic and erroneous strategy of over-predicting the most common class in the dataset. While this strategy allowed it to correctly identify 61% of all burglaries, it did so at the cost of incorrectly labeling a vast number of other crimes as burglaries. The confusion matrix visually corroborates this finding, showing that the model's predictions were overwhelmingly and incorrectly concentrated on the Burglary class, irrespective of the actual crime type. This behavior highlights a system that has not learned any meaningful patterns but has instead defaulted to a naive baseline, further cementing its lack of utility.

Explanation of Key Figures

This section provides a detailed explanation of the four key figures generated during the analysis, which collectively illustrate the model's performance, its internal logic, and the evidence of its encoded bias.

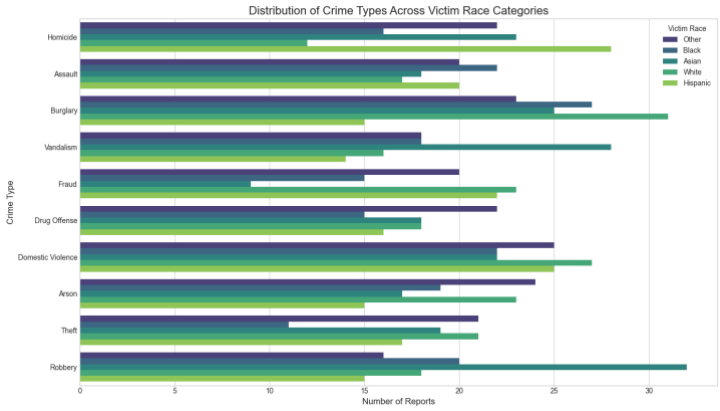

Figure 1 Distribution of Crime Types Across Victim Race Categories

Figure 1 is a product of the Exploratory Data Analysis (EDA) and serves as a critical baseline for the study. It visualizes the absolute frequency of each of the ten crime types, disaggregated by the race of the victim as recorded in the dataset. The primary purpose of this figure is to reveal the inherent statistical distributions and potential disparities present in the raw data before any predictive modeling is performed. For instance, it allows for a visual comparison of how often individuals from different racial groups are listed as victims for specific crimes like Burglary or Robbery. This initial view is crucial for contextualizing the model's subsequent behavior; if the model later relies on race to predict a certain crime, this chart helps to determine whether it is merely reflecting significant skew in the input data.

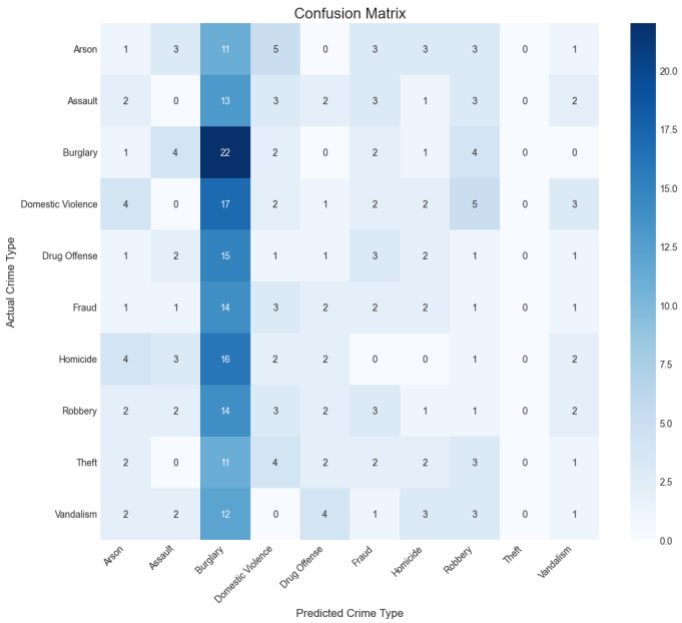

Figure 2 Confusion Matrix

Figure 2 is a direct visualization of the Decision Tree model's performance on the unseen test data. Each row in the matrix represents the instances in an actual class (the true crime type), while each column represents the instances in a predicted class. The values in the cells indicate the number of instances; for example, the cell at the intersection of the "Arson" row and the "Burglary" column shows how many actual Arson cases were incorrectly predicted as Burglary.

In an effective model, the highest values would be concentrated along the main diagonal (from top-left to bottom-right), indicating correct classifications. This matrix, however, shows the opposite. The values are scattered, and a large number of predictions are incorrectly concentrated in the "Burglary" column. This visually confirms the quantitative findings from the classification report: the model has extremely low accuracy and has defaulted to a naive strategy of over-predicting the most frequent crime type, failing to distinguish effectively between the different classes.

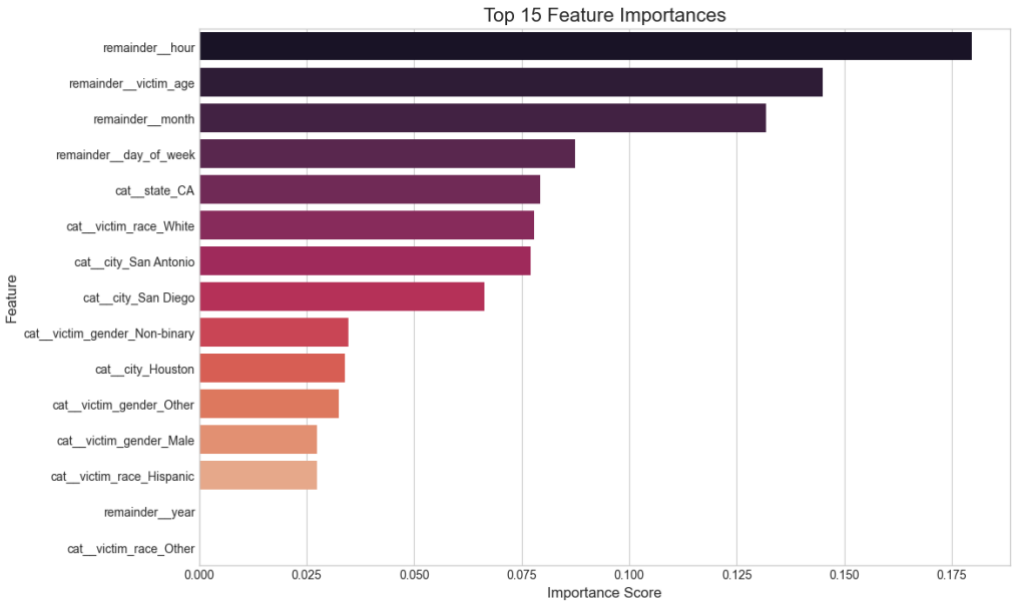

Figure 3 Top 15 Feature Importances

Figure 3 is one of the most critical pieces of evidence for the bias analysis, as it ranks the features that the Decision Tree model found most influential when making its predictions. The "importance score" on the x-axis is a metric calculated by the algorithm that quantifies how much each feature contributed to reducing uncertainty (or impurity) across the decision tree.

The key finding illustrated here is the high ranking of sensitive demographic attributes. While non-demographic features like remainder__hour and remainder__victim_age are ranked highest, protected characteristics such as cat__victim_race_White (6th), cat__victim_gender_Non-binary (9th), and cat__victim_race_Hispanic (13th) are also shown to be highly important. This chart provides direct, quantitative evidence that the model's predictive logic is significantly reliant on the victim's race and gender, which is the technical foundation for the legal argument of disparate impact.

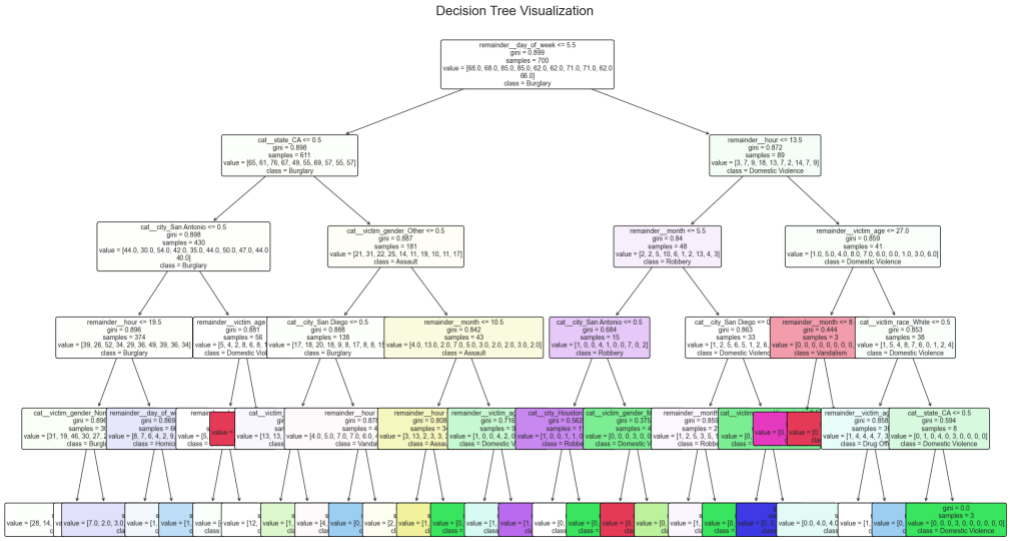

Figure 4 Decision Tree Visualization

Figure 4 provides an unprecedented level of transparency by mapping the exact decision-making logic of the trained model. Each node in the tree represents a "decision" based on a specific feature, splitting the data into branches based on the outcome of that decision (e.g., "is the day_of_week less than or equal to 5.5?"). By following the paths from the top (root) node down to the bottom (leaf) nodes, one can see the precise sequence of rules the model uses to arrive at a final classification for any given data point. This visualization serves as irrefutable proof of the model's discriminatory process. It moves beyond statistical correlation to show explicit, rule-based bias. For example, specific nodes in the tree show the model partitioning data based on rules like cat__victim_gender_Other <= 0.5 or using various racial categories to make further splits. This "white-box" view is the most direct evidence possible, clearly demonstrating that the algorithm has learned to use protected characteristics as fundamental components of its predictive logic.

Technical Interpretation of Results

Beyond the model's exceptionally poor predictive performance, an analysis of its internal logic provides the most critical insights, revealing not only that the model is biased, but that it may have resorted to discriminatory heuristics precisely because it failed to identify legitimate predictive patterns. The feature importance plot, which quantifies the relative influence of each variable on the model's classifications, serves as direct and unambiguous evidence of encoded bias. While temporal and age-related features (hour, victim_age, month) were ranked highest, sensitive demographic attributes were also identified as highly influential decision-making factors. Specifically, cat__victim_race_White emerged as the 6th most important feature, while cat__victim_gender_Non-binary and cat__victim_race_Hispanic ranked 9th and 13th, respectively. The model's significant reliance on these protected characteristics is a clear manifestation of algorithmic bias, wherein historical data disparities are learned and operationalized as predictive rules. The model has effectively determined that knowing a victim's race is one of the most salient factors for predicting crime type—a conclusion that is legally and ethically problematic because it codifies the principle that an individual's protected status can be used to infer criminality.

The Decision Tree visualization provides an even more explicit and granular confirmation of this bias. The visual representation of the model's decision-making process shows clear, specific instances where the algorithm explicitly partitions the data based on protected characteristics. For example, a prominent node near the top of the tree utilizes the rule cat__victim_gender_Other <= 0.5 to partition hundreds of data points, directly incorporating a victim's gender identity into its core predictive logic. Deeper within the tree, other decision nodes are predicated on racial categories, creating divergent analytical paths for individuals based on their demographic profile. The "white-box" nature of the Decision Tree eliminates ambiguity; this is not a matter of subtle correlation, but a clear, hard-coded demonstration that the model learned to use race and gender as primary factors in its classification rules. This technical finding is the fulcrum of the legal analysis, as it provides interpretable and irrefutable evidence of the model's discriminatory logic, making it impossible to argue that the bias is an unintended or opaque side effect.

Legal and Policy Implications of Findings

The technical results have profound and troubling implications from a cyber law and constitutional perspective. First, the deployment of a system with a 10% accuracy rate would likely constitute a significant deviation from the standard of care required of a state actor. Utilizing such a fundamentally unreliable tool to inform the allocation of public resources or to justify police actions could be deemed arbitrary and capricious. The model is not merely a biased tool; it is a fundamentally flawed one, and its application in the field would foreseeably lead to inefficient, ineffective, and unjust outcomes.

Second, the model's demonstrated reliance on protected characteristics creates a clear and actionable case of "disparate impact" under the Equal Protection Clause of the Fourteenth Amendment and various civil rights statutes. The feature importance plot and the Decision Tree visualization would serve as compelling evidence in litigation, proving that the system, even if not designed with discriminatory intent, produces discriminatory outcomes by treating individuals differently based on their race and gender. The interpretability of the Decision Tree is a critical factor; unlike a "black-box" model where bias must be inferred from outcomes alone, this model's logic is transparently discriminatory, rendering any claim of ignorance on the part of the deploying agency untenable. For a municipality, deploying this system would create substantial legal and financial liabilities, as the feature importance plot alone could be presented as primary evidence in a lawsuit alleging systemic algorithmic discrimination.

Furthermore, any police action influenced by this model's output would be of questionable constitutional validity. The model's predictions could not plausibly contribute to the "totality of the circumstances" required for establishing probable cause or reasonable suspicion, as they are derived from a process that is both demonstrably biased and overwhelmingly inaccurate. Evidence gathered as a result of an investigation prompted by this model could be challenged in court as the "fruit of a poisonous tree," with the poisonous source being the unconstitutional, discriminatory algorithm.

Comparison with Previous Research

The findings of this analysis are consistent with and contribute to a growing body of literature that critically examines algorithmic systems in the criminal justice sector. The identification of racial bias aligns with seminal investigative work, such as ProPublica's analysis of the COMPAS recidivism algorithm, which found that the tool was more likely to falsely flag Black defendants as future criminals. However, this study extends that line of inquiry by utilizing an interpretable "white-box" model. While much of the existing research has focused on demonstrating the disparate outcomes of opaque, "black-box" systems, the use of a Decision Tree in this analysis allows for the direct inspection of the discriminatory logic itself. This provides a more direct and arguably more powerful form of evidence for legal review, moving the debate from statistical inference to a direct examination of the algorithm's decision-making rules.

Limitations of the Study

It is important to acknowledge the limitations of this research. The primary limitation is the nature of the dataset; the analysis was conducted on a simulated, publicly available dataset rather than on real-world, operational data from a specific police department. As such, the findings demonstrate a proof-of-concept for the auditing methodology rather than an indictment of any specific system currently in use. Secondly, the dataset's relatively small size (1,000

records) may constrain the model's ability to learn complex, non-linear patterns, potentially amplifying its reliance on simplistic and biased heuristics. Finally, this study focused on a single classification algorithm. While the Decision Tree was chosen for its interpretability, other machine learning architectures might produce different results, although they would be subject to the same underlying biases present in the data.

Future Research Suggestions

Building on this analysis, several avenues for future research are recommended. First, there is a critical need for studies that apply this auditing methodology to real-world predictive policing systems, which would require data-sharing partnerships between academic researchers and municipal agencies. Such collaborations are essential for moving from theoretical analysis to practical oversight. Second, future interdisciplinary research should focus on the development of "fairness-aware" machine learning techniques. This would involve not only the technical implementation of bias-mitigation strategies during model training but also a rigorous legal evaluation of whether these technical "fixes" are sufficient to meet constitutional standards of due process and equal protection. Finally, further research should investigate the downstream, real-world impacts of these systems, exploring how potentially biased algorithmic recommendations influence police officer behavior and whether they create self-perpetuating feedback loops that exacerbate existing inequalities in the criminal justice system.

Conclusion

This study demonstrated the profound risks inherent in the unaudited application of machine learning to predictive policing. Through the construction and analysis of a Decision Tree classifier, it was revealed that the model was not only functionally ineffective, achieving a mere 10% accuracy, but was also fundamentally biased. The model's internal logic, made transparent by the interpretability of the Decision Tree, showed a clear and systematic reliance on protected demographic characteristics—including race and gender—to make its predictions. This dual failure of performance and fairness illustrates a critical disconnect between the technical implementation of such systems and the legal and ethical standards required for their use in the public sector.

Ultimately, this research serves as a cautionary proof-of-concept, highlighting that without rigorous oversight, predictive policing technologies risk becoming instruments of injustice rather than tools of public safety. The findings underscore the urgent need for robust regulatory frameworks that mandate independent, transparent audits of any algorithmic system intended for use by government agencies. Such audits must include a detailed analysis of feature importances and decision pathways to expose and mitigate encoded biases before deployment. Moving forward, a collaborative, interdisciplinary approach between data scientists, legal scholars, and policymakers is essential to ensure that the pursuit of technological innovation in law enforcement does not come at the cost of fundamental constitutional rights.

Declarations

Author Contributions

Conceptualization: L.Y.; Methodology: M.P.; Software: L.Y.; Validation: M.P.;

Formal Analysis: M.P.; Investigation: L.Y.; Resources: M.P.; Data Curation: L.Y.; Writing Original Draft Preparation: M.P.; Writing Review and Editing: M.P.; Visualization: L.Y.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Khan and M. Khan, “Real-Time Data Processing Systems in Modern Law Enforcement: A Technical Analysis,” *Int. J. Multidiscip. Res.*, 2024, doi: 10.36948/ijfmr.2024.v06i06.30723.

[2] A. Sandhu and P. Fussey, “The ‘Uberization of Policing’? How Police Negotiate and Operationalise Predictive Policing Technology,” *Polic. Soc.*, 2020, doi: 10.1080/10439463.2020.1803315.

[3] D. Gerstner, “Predictive Policing in the Context of Residential Burglary: An Empirical Illustration on the Basis of a Pilot Project in Baden-Württemberg, Germany,” *Eur. J. Secur. Res.*, 2018, doi: 10.1007/s41125-018-0033-0.

[4] D. Wilson, “Predictive Policing Management: A Brief History of Patrol Automation,” *New Form.*, 2019, doi: 10.3898/newf:98.09.2019.

[5] B. Benbouzid, “To Predict and to Manage. Predictive Policing in the United States,” *Big Data Soc.*, 2019, doi: 10.1177/2053951719861703.

[6] L. Strikwerda, “Predictive Policing: The Risks Associated With Risk Assessment,” *Police J. Theory Pract. Princ.*, 2020, doi: 10.1177/0032258x20947749.

[7] K. Lum and W. Isaac, “To Predict and Serve?,” *Significance*, 2016, doi: 10.1111/j.1740-9713.2016.00960.x.

[8] S. Brayne, “The Criminal Law and Law Enforcement Implications of Big Data,” *Annu. Rev. Law Soc. Sci.*, 2018, doi: 10.1146/annurev-lawsocsci-101317-030839.

[9] A. P. Hutama, A. J. Simon Runturambi, and A. Iskandar, “The RW Police Program as an Implementation of Predictive Policing in the Legal Jurisdiction of the Jakarta Metropolitan Police Department (Polda Metro Jaya),” *Int. J. Multicult. Multireligious Underst.*, 2023, doi: 10.18415/ijmmu.v10i5.4756.

[10] M. Oswald, J. Grace, S. Urwin, and G. C. Barnes, “Algorithmic Risk Assessment Policing Models: Lessons From the Durham HART Model and ‘Experimental’ Proportionality,” *Inf. Commun. Technol. Law*, 2018, doi: 10.1080/13600834.2018.1458455.

[11] A. K. Suud, “The Building Public Trust Against to Law Enforcers in the Covid 19

Pandemic," *Int. J. Res. Community Serv.*, 2022, doi: 10.46336/ijrcs.v3i1.183.
[12] A. Wheeler and W. Steenbeek, "Mapping the Risk Terrain for Crime Using Machine Learning," *J. Quant. Criminol.*, 2020, doi: 10.1007/s10940-020-09457-7.
[13] G. Mohler and M. D. Porter, "Rotational Grid, PAI-maximizing Crime Forecasts," *Stat. Anal. Data Min. Asa Data Sci. J.*, 2018, doi: 10.1002/sam.11389.
[14] S. Garnier, J. M. Caplan, and L. W. Kennedy, "Predicting Dynamical Crime Distribution From Environmental and Social Influences," *Front. Appl. Math. Stat.*, 2018, doi: 10.3389/fams.2018.00013.
[15] P. Cerda and G. Varoquaux, "Encoding High-Cardinality String Categorical Variables," *Ieee Trans. Knowl. Data Eng.*, 2022, doi: 10.1109/tkde.2020.2992529.
[16] J. Adams-Prassl, R. Binns, and A. Kelly-Lyth, "Directly Discriminatory Algorithms," *Mod. Law Rev.*, 2022, doi: 10.1111/1468-2230.12759.
[17] M. E. Gilman, "Expanding Civil Rights to Combat Digital Discrimination on the Basis of Poverty," *Smu Law Rev.*, 2022, doi: 10.25172/smulr.75.3.6.
[18] Д. Базарова, "Experience of Developed Countries in the Application of Artificial Intelligence to Ensure the Protection of Personal Rights in Criminal Justice," *Rev. Law Sci.*, 2022, doi: 10.51788/tsul.rols.2022.6.1./irbs2461.
[19] J. S. Rathore, "Legal and Ethical Implications of Predictive Policing Technologies," *J. Adv. Sch. Res. Allied Educ.*, 2019, doi: 10.29070/p2qqng11.
[20] D. J. del Bueno, "Gender Bias in Artificial Intelligence: A Critical Perspective and Legal Analysis," *Ac*, 2024, doi: 10.22201/fder.23959045e.2024.26.90464.
[21] R. K. Sharma, "Ethics in AI: Balancing Innovation and Responsibility," *Int. J. Sci. Res. Arch.*, 2025, doi: 10.30574/ijsra.2025.14.1.0122.