Use of Feature Blindness To Improve Recidivism Predictions

Daniel Joseph Miller[#1], Shinyoung Kim[*2]

[#]*College of Information and Science, Pennsylvania State University*

[1]djm7829@psu.edu

[2]sjk6402@psu.edu

ABSTRACT

The criminal justice system faces significant challenges in addressing misdemeanor recidivism, particularly concerning fairness and bias in predictive modeling. This paper explores the concept of predictive fairness in recidivism predictions, emphasizing the integration of social service interventions to mitigate risks. We evaluate various machine learning models, including decision trees and random forests, to assess their predictive accuracy and fairness metrics. Our findings highlight the necessity for equitable treatment across demographic groups and the importance of collaboration between data science and public policy to enhance recidivism reduction efforts. Ultimately, we envision a justice system where technology supports rehabilitation while minimizing bias.

KEYWORDS

Recidivism, Artificial intelligence, Machine Learning, Bias, Discrimination, Mitigation-Stragies, Fairness, COMPAS

I. INTRODUCTION

The increasing prevalence of predictive analytics in the criminal justice system has raised critical ethical concerns, particularly regarding fairness and bias. As jurisdictions across the globe adopt data-driven approaches to inform decisions about sentencing, granting or denying parole, and even rehabilitation programs; the potential for these systems to perpetuate existing inequalities becomes a significant issue. The integration of machine learning and artificial intelligence into criminal justice processes promises efficiency and objectivity; however, it also poses risks of entrenching systemic biases that disproportionately affect marginalized communities.

Misdemeanor recidivism remains a pressing concern, with many individuals cycling through the system due to inadequate support and intervention strategies. Recidivism, particularly for minor offenses, is not merely a reflection of individual behavior but is deeply intertwined with social determinants such as poverty, education, and community resources. Traditional predictive models often fail to account for these complexities, leading to oversimplified assessments that overlook the multifaceted nature of human behavior. Consequently, these models can result in disproportionate impacts on marginalized communities, where individuals may be unfairly categorized as high-risk based on historical data that reflects systemic inequalities rather than their actual likelihood of reoffending.

The ethical implications of using such predictive models are profound. For instance, individuals from underrepresented racial and ethnic backgrounds may be subjected to harsher penalties or denied opportunities for rehabilitation based on flawed algorithms that do not consider the broader context of their circumstances. One specific program that was implemented was that of the Recidivism Reduction and Drug Diversion Unit (R2D2), a program utilizing data analytics to allocate their resources in an effort to reduce misdemeanor recidivism across Los Angeles [1]. Amongst many other programs, this raises critical questions about the fairness of a system that relies on data that may be biased, as well as the accountability of those who design and implement these predictive tools. The potential for algorithmic bias to influence life-altering decisions underscores the urgent need for a more equitable approach to predictive analytics in criminal justice.

Moreover, the punitive nature of the criminal justice system often overshadows the need for rehabilitative approaches. Many individuals who reoffend do so not out of a lack of will but due to a lack of resources and support to acclimate into society successfully following a time being incarcerated. The current system frequently prioritizes punishment over rehabilitation, which can exacerbate the challenges faced by individuals upon their release. This highlights the urgent need for a paradigm shift in how we approach recidivism, moving from a purely punitive framework to one that emphasizes support, intervention, and rehabilitation.

This paper aims to address these challenges by proposing an innovation upon previously created comprehensive frameworks that integrate predictive fairness with social service interventions. By leveraging advanced machine learning techniques, we seek to develop a model that not only predicts recidivism risk but also promotes equitable treatment and rehabilitation opportunities for all individuals. Our goal is to create a system that recognizes the unique circumstances of everyone, enabling for tailored interventions that can effectively reduce recidivism rates while fostering a more just and equitable criminal justice system.

II. LITERACTURE REVIEW

A. Existing Research on Predictive Models

Predictive models have increasingly been utilized in the criminal justice system to assess the likelihood of recidivism and inform decision-making processes. These models often leverage historical data to identify patterns and predict future behaviors, aiming to enhance the efficiency of resource allocation and intervention strategies. Research has shown that various machine learning techniques, such as logistic regression, decision trees, and random forests, can effectively predict recidivism rates, providing valuable insights for law enforcement and judicial systems. For instance, researchers from the West Ukrainian National University were able to achieve notable results within

criminal recidivism predictions with models that include Naïve Bayes, Generalized Linear, Logistic Regression, and many more [2]. Properly preparing data and training these models can identify high-risk individuals with very low error, allowing for targeted interventions that may prevent future offenses. However, the effectiveness of these models is contingent upon the quality and representativeness of the data used, as well as the algorithms employed.

Taking precautions to ensure for data integrity is absolutely critical to the assurance of quality results; accounting for fairness within the handling of the data throughout the machine learning process is integral to mitigating the possible sources of bias within the models themselves. Amongst many methods that can be used to remove bias from the model creation process, the primary ideas that are discussed within modern research are pre-processing, model selection and post-processing decisions. Practically speaking, these principles can take shape in many ways, such as under or oversampling the original data to address a unbalanced feature or creating more data to balance the features. For an example of post processing, performing a post-hoc adjustment to balance the results can remove a systematic bias that may be present due to an imbalance in one or more features present [1].

Overall, all of these ideas have heavily influenced the methods that we have employed within our methodology to process data and create machine learning models, mitigating the potential bias that may be present in our results.

B. Traditional COMPAS Models and Their Limitations

The Traditional COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) models are algorithms used in the criminal justice system to assess the likelihood of a defendant reoffending. These models generate risk scores based on various factors, including criminal history, demographic information, and other relevant data. However, there are several limitations associated with these models. The models have been criticized for exhibiting racial bias. Studies have shown that the algorithm tends to assign higher risk scores to Black defendants compared to white defendants, even when controlling for prior criminal behavior. This raises concerns about fairness and the potential for perpetuating systemic inequalities in the justice system.

The effectiveness of COMPAS models heavily relies on the quality and completeness of the input data. Incomplete or inaccurate data can lead to misleading risk assessments. For instance, if certain demographic groups are underrepresented in the dataset, the model may not accurately reflect their recidivism risk.

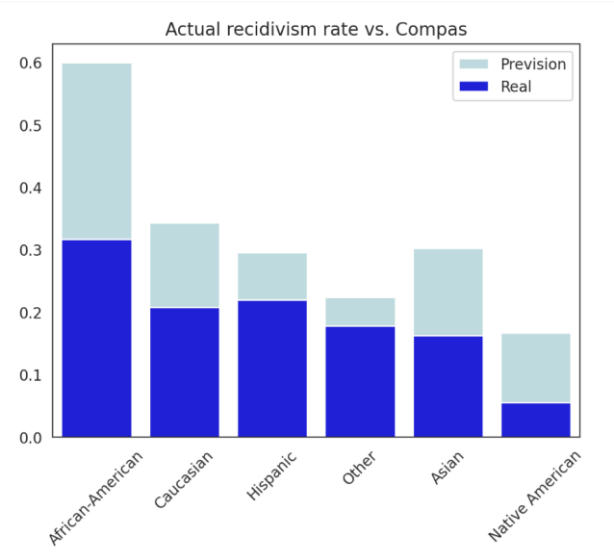

Fig. 1. The actual recidivism rate vs. that predicted by COMPAS

Analyzing the COMPAS score predictions has been performed extensively, with the system receiving high levels of scrutiny due to accusations of unbalanced false negatives from individuals of different races. This is shown in Fig. 1 and Fig. 2, which shows the difference between the actual recidivism rates and those of the predictions performed by ProPublica [4]. Regardless of the correctness of the accusations, this denoted an incredible driving force within a changing view surrounding the widespread use of COMPAS and raising the emphasis of machine learning bias for the creation of future recidivism prediction models.

	Native American	FPR	FNR
0	Other	0.177778	0.564103
1	African-American	0.534973	0.261806
2	Caucasian	0.284190	0.431481
3	Hispanic	0.217973	0.428571
4	Asian	0.194444	0.142857
5	Native American	0.117647	0.000000

Fig. 2. False Positive and False Negative score by race

C. Ethical Concerns and Biases

The use of predictive analytics in criminal justice has sparked significant ethical debates, particularly regarding issues of bias and fairness. Studies have demonstrated that predictive models can perpetuate existing biases present in historical data, leading to disproportionate impacts on marginalized communities. For instance, racial and gender biases can be inadvertently encoded into algorithms, resulting in unfair treatment and reinforcing systemic inequalities. Research has shown that individuals from minority backgrounds may be overrepresented in datasets used to train predictive models, leading to skewed predictions that do not accurately reflect their risk levels. As such, it is crucial to critically examine the ethical implications of these models and implement strategies to mitigate bias and enhance fairness. This includes not only refining the algorithms themselves but also ensuring that the data used is representative and free from historical biases.

D. Recidivism Studies and Tailored Interventions

As with applying machine learning to various domains, a one-size-fits-all approach to recidivism prediction is insufficient. Studies have highlighted the importance of tailored interventions that consider individual circumstances and needs. By integrating social services and support systems into predictive models, practitioners can develop more effective strategies for rehabilitation. This approach not only addresses the root causes of criminal behavior but also promotes equity by ensuring that interventions are responsive to the diverse needs of individuals within the justice system. For example, individuals with mental health issues may require different support than those with substance abuse problems. By recognizing these differences, predictive models can facilitate more personalized interventions that are more likely to succeed in reducing recidivism rates. This is the reasoning behind the aforementioned R2D2 intervention program suggested by the Los Angeles City Attorney's Office [1]. However, this idea of coordination between data analytics and domain professionals is not limited to this program, as it is crucial as a president within various applications in data science.

III. Research Motivation

A. Need for Equity in Predictive Models

The need for equity in predictive models, particularly in the context of recidivism risk assessment, is increasingly recognized as a critical issue in the criminal justice system. One of the primary reasons for this need is to address bias. Predictive models, such as COMPAS, have been shown to exhibit biases against certain demographic groups, particularly racial minorities. This can lead to disproportionate risk assessments that unfairly impact individuals based on their race or ethnicity. Ensuring equity in predictive models is essential to mitigate these biases and promote fair treatment in the justice system. Fairness in decision-making is another crucial aspect of equity in predictive models. It is vital to ensure that decisions made based on these models—such as sentencing, parole, and probation—are fair and just. When models are biased, they can perpetuate systemic inequalities and lead to unjust outcomes for marginalized communities. Striving for equity helps to uphold the principles of justice and fairness in legal proceedings.

B. Addressing Inaccuracies and Biases

Addressing inaccuracies and biases in predictive models is essential for fostering trust in the criminal justice system. Racial and gender disparities in recidivism predictions can lead to unjust outcomes, including wrongful incarceration and stigmatization of certain groups. By actively working to identify and rectify these biases, researchers can enhance the validity of predictive models and promote a more equitable approach to justice. This involves not only refining the algorithms but also engaging with community stakeholders to understand their perspectives and experiences. By incorporating feedback from those directly affected by the justice system, researchers can develop models that are more reflective of the realities faced by individuals in these communities.

C. Significance of Tailored Interventions

Tailored interventions are critical for effectively reducing recidivism rates. By recognizing that individuals have unique needs and circumstances, practitioners can design interventions that are more likely to succeed. This personalized approach not only improves outcomes for individuals but also contributes to the overall effectiveness of the criminal justice system. Integrating social services and support mechanisms into predictive models can facilitate early interventions, ultimately leading to a reduction in recidivism and a more rehabilitative justice system. Furthermore, by focusing on rehabilitation rather than punishment, the justice system can promote positive outcomes for individuals and communities, fostering a more just and equitable society.

IV. Methodology

A. Data Used

The data that we used was gathered by ProPublica and published on Kaggle, to share the analysis that they conducted of COMPAS and encourage others to verify their findings. This data set contained various binary and categorical features along with the recidivism feature. Amongst these features are the number of previous juvenile felonies, age, race, sex, how long they were in custody, number of days since release, and more. The shortcoming with this dataset is almost certainly the number of rows, with only 18,316 cases included [4].

B. Machine Learning Techniques

This study employs a range of machine learning techniques to develop predictive models for recidivism. Techniques such as decision trees and random forests will be utilized to analyze historical data and identify key predictors of recidivism. These methods allow for the exploration of complex relationships within the data and provide a robust framework for making predictions. Additionally, advanced techniques such as support vector machines and neural networks may be considered to further enhance predictive accuracy, but we have opted to focus on the decision tree and random forest models to emphasize the bias reduction technique performed. Ultimately, the choice of algorithms will be guided by their performance in terms of accuracy, interpretability, and fairness.

The primary reason for choosing to work with decision trees was the formatting of the data. With the majority of the features as binary variables, decision trees are ideal to narrow down the result. This has also been applied within the field of recidivism prediction, with results detailing recall and precision over 97% [2].

C. Integrating Predictive Fairness

The input dataset used in this study exhibited a significant imbalance, with one race predominating over others. This imbalance could introduce biases in model predictions, as groups with fewer samples may not be adequately represented during training. To address this issue and ensure fairness in predictions, we applied various preprocessing techniques to mitigate potential biases arising from the dataset's structure.

race	count
African-American	5890
Caucasian	3664
Hispanic	835
Other	626
Asian	47
Native American	23

dtype: int64

Fig. 3. Balance of Race by COMPAS dataset

sex	count
Male	5837
Female	1491

dtype: int64

Fig. 4. Balance of Sex by COMPAS dataset

To reduce the influence of disproportionate representation in the dataset, we balanced the data by race and sex. Specifically, we adjusted the sample sizes for the two most prevalent race categories, African American and Caucasian, to achieve parity. The African American class was down sampled to match the number of samples in the Caucasian class. Similarly, the dataset was balanced by sex by aligning the number of male samples to the female samples' count. These steps ensured equal representation of the major subgroups, reducing the risk of predictive bias driven by data prevalence. This achieved the desired goal, but at the cost of some data which could result in information loss, resulting in a worse performance of the model.

The balanced dataset was constructed as follows:

- African American and Caucasian records were sampled to achieve equal class sizes.

race	count
African-American	3664
Caucasian	3664
Hispanic	0
Other	0
Asian	0
Native American	0

dtype: int64

Fig. 5. Balanced dataset achieved by equal class sizes of African American and Caucasian

- Male and female records were then balanced within the already race-balanced dataset.

```
<class 'pandas.core.frame.DataFrame'>
Index: 2982 entries, 3513 to 18303
Data columns (total 16 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   id                       1737 non-null   float64
 1   sex                      2982 non-null   category
 2   age                      2982 non-null   int64
 3   race                     2982 non-null   category
 4   c_charge_degree          2982 non-null   category
 5   c_charge_desc            2979 non-null   object
 6   days_b_screening_arrest  2982 non-null   float64
 7   decile_score             2982 non-null   int64
 8   is_recid                 2982 non-null   int64
 9   c_case_number            2982 non-null   object
 10  v_decile_score           2982 non-null   int64
 11  score_text               2982 non-null   category
 12  days                     82 non-null     float64
 13  two                      2982 non-null   int64
 14  two_years_r              2982 non-null   int64
 15  two_years_v              2982 non-null   int64
dtypes: category(4), float64(3), int64(7), object(2)
memory usage: 315.6+ KB
```

Fig. 6. Balanced dataset achieved by equal class sizes of Male and Female

These balancing steps resulted in a dataset that is better suited for evaluating predictive fairness across demographic groups, enabling the model to focus on substantive relationships rather than being influenced by overrepresented subgroups. In addition to balancing the dataset, we conducted feature engineering to further minimize the influence of sensitive attributes, such as race and sex on predictions.

D. Analysis Methods

Data preprocessing steps will be undertaken to clean and prepare the data for analysis, including handling missing values, normalizing data, and encoding categorical variables. Statistical techniques used to compare the impact of the data pre-processing included precision, f1-score, Precision, and ROCAUC curve. The study will employ cross-validation techniques throughout different decision trees to assess model performance and generalizability in the form of a random forest classifier, ensuring that the findings are robust and applicable to real-world settings.

We conducted an EDA to gain insights into the dataset's structure and identify potential biases. This includes visualizing the distribution of features, examining correlations, and identifying outliers. Techniques such as histograms, box plots, and scatter plots are employed to illustrate findings. Statistical tests, such as chi-square tests for categorical variables and t-tests for continuous variables, are performed to evaluate the significance of differences in recidivism rates across demographic groups. This helps to quantify the extent of bias present in the COMPAS algorithm. Various visualization techniques are employed to present findings clearly and effectively. These visuals are used to illustrate disparities in recidivism rates and model predictions across different demographic groups.

E. Evaluation and Validation

The proposed preprocessing framework was rigorously validated to assess the impact of removing sensitive attributes (race and sex) on the fairness and predictive performance of machine learning models. The validation process involved training and testing multiple models on three distinct versions of the dataset—one excluding race (df_final_noR), one excluding sex (df_final_noS), and one excluding both race and sex (df_final_noRS). This allowed for a comprehensive evaluation of how omitting sensitive features affects both model accuracy and fairness.

F. Data Preprocessing

The initial step involves loading the dataset from the ProPublica COMPAS analysis repository. The dataset contains various features, including demographic information, criminal history, and COMPAS risk scores. Data cleaning is crucial to ensure the integrity of the analysis. This includes handling missing values, correcting data types, and removing duplicates. Several columns in the dataset represent dates (e.g., birth dates, arrest dates), which are parsed into appropriate datetime formats using pandas, facilitating temporal analysis and ensuring accurate calculations of age and time served.

The dataset is often imbalanced, particularly concerning recidivism rates among different demographic groups, which can introduce biases in predictions. To address this, we incorporated a feature blindness approach in the preprocessing pipeline to mitigate potential biases arising from sensitive attributes. Specifically, we removed the race and sex attributes, either individually or together, to create three versions of the dataset: one without race (df_final_noR), one without sex (df_final_noS), and another excluding both (df_final_noRS). By omitting these sensitive features, the analysis focuses on other predictors, such as age, criminal background, years in prison, and type of crime, which are more directly linked to the prediction target. This approach ensures that the model's predictions are not directly influenced by demographic information, promoting fairness in the analysis while enabling an evaluation of the extent to which these attributes impact model performance.

Specifically, we created three variations of the dataset with sensitive attributes removed:

- No Race Dataset: The race feature was removed.
- No Sex Dataset: The sex feature was removed.
- No Race and Sex Dataset: Both the race and sex features were removed

These variations allowed us to examine the impact of excluding sensitive features on the predictive performance and fairness of the model. By removing these features, the model was encouraged to rely on other predictors, such as age, criminal background, years in prison, and type of crime, which are more directly related to the target variable.

G. Model Selection

For this study, we chose Random Forest and Decision Tree models for their interpretability, ease of implementation, and strong performance in classification tasks. Both models were selected because they provide a transparent understanding of how different features contribute to the prediction, which is essential in the context of criminal justice risk assessments where the interpretability of decisions is crucial.

Decision Tree: The Decision Tree algorithm was selected for its simplicity and clarity. It offers a straightforward approach to classification by recursively splitting the data into subsets based on feature values, leading to a tree-like structure. The model's transparency allows for easy interpretation of how input features, such as criminal history, age, and type of crime, contribute to the prediction. Although decision trees are prone to overfitting, they provide a solid baseline model for understanding basic relationships in the data.

```
For Model dt_model_balanced, Accuracy: 0.640, f1-score: 0.559, ROC_AUC: 0.631, Precision: 0.654
For Model dt_model_noR, Accuracy: 0.635, f1-score: 0.625, ROC_AUC: 0.635, Precision: 0.636
For Model dt_model_noS, Accuracy: 0.638, f1-score: 0.642, ROC_AUC: 0.638, Precision: 0.629
For Model dt_model_noRS, Accuracy: 0.631, f1-score: 0.636, ROC_AUC: 0.631, Precision: 0.621
```

Fig. 7. Evaluation of Decision Tree model

Random Forest: To address the overfitting limitations of individual decision trees, we used Random Forest, an ensemble learning method that builds multiple decision trees and combines their outputs for improved performance and stability. Random Forest typically offers better generalization by averaging predictions from several trees, which helps reduce model variance. This approach is particularly valuable when working with complex datasets like COMPAS, where relationships between features can be intricate and noisy. Random Forest also handles missing values well and is less sensitive to outliers, making it a robust choice for real-world datasets.

```
For Model rf_model_balanced, Accuracy: 0.638, f1-score: 0.563, ROC_AUC: 0.630, Precision: 0.647
For Model rf_model_noR, Accuracy: 0.635, f1-score: 0.625, ROC_AUC: 0.635, Precision: 0.636
For Model rf_model_noS, Accuracy: 0.636, f1-score: 0.641, ROC_AUC: 0.637, Precision: 0.627
For Model rf_model_noRS, Accuracy: 0.631, f1-score: 0.636, ROC_AUC: 0.631, Precision: 0.621
```

Fig. 8. Evaluation of Random Forest model

Both models were trained on three versions of the dataset: one with all features, and two versions with sensitive features (race and sex) removed. The models' performance was evaluated based on standard classification metrics (accuracy, precision, recall, F1 score) and fairness metrics (demographic parity, equal opportunity). By comparing these models, we were able to assess how the exclusion of sensitive attributes impacts predictive accuracy and fairness.

The results revealed a clear trade-off between fairness and predictive performance. Models trained on datasets excluding race or sex showed improvements in fairness, with reductions in the demographic parity and equal opportunity differences. However, they experienced slight declines in predictive accuracy compared to models trained with all features. This demonstrates the inherent challenge of balancing equity and performance, especially in sensitive applications like criminal justice risk assessments. Despite this trade-off, the evaluation confirmed that our approach contributed to mitigating potential biases, ensuring that demographic information did not directly influence predictions.

The evaluation revealed several key insights. Models trained on the full dataset, which included sensitive features like race and sex, generally achieved higher accuracy. However, this came at the cost of fairness, as predictions for minority groups showed disproportionate rates of false positives and false negatives. This underscores the ethical concerns surrounding the inclusion of sensitive attributes in predictive models. The primary goal of the fairness analysis was to assess how well the models could predict recidivism without relying on biased demographic features. Removing race and sex from the dataset led to more equitable outcomes. Specifically, models trained on the datasets df_final_noR (without race) and df_final_noS (without sex) showed improvements in fairness. The model trained on df_final_noRS (without both race and sex) exhibited the most significant fairness gains, reducing both the demographic parity difference and the equal opportunity difference.

However, a thorough evaluation of model biases revealed that certain groups, particularly racial minorities, still received disproportionately high or low risk scores, even after excluding sensitive features. This suggests that race-related biases may still be captured through non-race-related features. While the feature blindness approach improved fairness, some predictive bias persisted, especially in the model trained on df_final_noR. The models were tested on

different data splits, with results averaged to minimize the impact of specific data partitions. This helped evaluate the stability of the models across various datasets and their ability to avoid overfitting.

V. MODEL RESULTS

Overall, the evaluation underscores the ongoing challenges of balancing fairness and predictive performance in criminal justice risk assessments. This can be seen within Fig. 7-8, showing how the different training data impacted the fairness evaluation metrics. Furthermore, While removing sensitive attributes significantly improved fairness, it came at the cost of slight declines in predictive performance. These findings highlight the importance of exploring more sophisticated bias mitigation strategies in future work to further address these challenges and ensure that machine learning models can deliver both fairness and accuracy.

The failure to achieve results that breach a value of 0.7 for the ROCAUC curve is likely reflective of the number of data that was removed in order to achieve our goals of pre-processing the data. However, the trends shown within the data shows that blinding the data does not result in significant information loss relative to the accuracy or precision of the predictions achieved.

VI. CONCLUSION

This research underscores the critical importance of addressing bias in predictive models within the criminal justice system, particularly in the context of risk assessments such as the false negative rate by sex. By removing sensitive demographic attributes such as race and sex, we were able to mitigate some of the direct biases embedded in the dataset. The results demonstrate that the omission of these attributes leads to improved fairness, as measured by a reduction in demographic parity and equal opportunity differences. However, it is also evident that there is a trade-off between fairness and accuracy, with the models trained on datasets excluding sensitive attributes showing slightly lower predictive accuracy compared to those trained on the full dataset.

Our findings emphasize that fairness-aware machine learning models is necessary to ensure equitable outcomes in high-stakes applications like recidivism. At the same time, they also reveal that achieving fairness requires careful consideration of both direct and indirect biases. Future work will need to further explore advanced debiasing techniques and their integration into machine learning models to create ethical, transparent, and effective systems.

VII. DISCUSSION AND FUTURE WORK

The findings of this study hold significant implications for the use of machine learning in criminal justice risk assessments. By removing sensitive attributes such as race and sex, we were able to reduce some of the bias inherent in the recidivism dataset. However, this approach did not eliminate all forms of bias, particularly those that may be indirectly captured through other features. The persistence of predictive bias, even after excluding demographic information, highlights the need for more advanced debiasing techniques and a deeper understanding of how systemic issues might be reflected in the data.

Moreover, the trade-off between fairness and accuracy remains a key challenge. While fairness-aware preprocessing improved equity by reducing demographic bias, it did so at the expense of predictive performance. This trade-off underscores the difficult balancing act between optimizing predictive models and ensuring fairness, a critical consideration when the outcomes of these models can directly affect individuals' lives. The results call for ongoing research into methodologies that can enhance both fairness and predictive power, with special attention to the social and ethical consequences of deploying such models in practice.

This study also raises broader questions about the use of predictive algorithms in criminal justice. While machine learning models have the potential to improve decision-making and reduce human bias, they may also reinforce historical inequalities if not properly managed. Therefore, ethical considerations must guide the development and deployment of these technologies, ensuring that they contribute to the larger goals of fairness and justice.

Several avenues for future research could build upon the findings of this study. While the feature blindness approach effectively reduced direct biases, there remains a need to explore additional bias mitigation strategies, such as adversarial debiasing or causal inference models, which may help address indirect biases that persist within the data. Incorporating longitudinal data, which tracks individuals over time, could provide valuable insights into the long-term effectiveness of predictive models, offering the opportunity to assess how the fairness and predictive performance of these models evolve as individuals' circumstances change.

Furthermore, future studies should prioritize the real-world deployment of these models, working closely with criminal justice professionals to evaluate their practical application in predicting recidivism and determining parole eligibility. Engaging stakeholders and incorporating real-time feedback could help refine model predictions, ensuring they align with the practical needs and challenges faced by the criminal justice system.T

REFERENCE

[1] Rodolfa, K. T., Salomon, E., Haynes, L., Mendieta, I. H., Larson, J., Ghani, R., Kit T. RodolfaCarnegie Mellon UniversityView Profile, Profile, E. S. of C., Profile, L. H. of C., Iván Higuera MendietaUniversity of ChicagoView Profile, Jamie LarsonLos Angeles City Attorney's OfficeView Profile, & Profile, R. G. M. U. (2020, January 27). *Case study: Proceedings of the 2020 conference on fairness, accountability, and transparency*. ACM Conferences. https://dl.acm.org/doi/10.1145/3351095.3372863

[2] Kovalchuk, O., Karpinski, M., Banakh, S., Kasianchuk, M., Shevchuk, R., & Zagorodna, N. (2023, March 3). *Prediction machine learning models on propensity convicts to criminal recidivism*. MDPI. https://doi.org/10.3390/info14030161

[3] Ferrara, E. (2023, December 26). *Fairness and bias in Artificial Intelligence: A brief survey of sources, impacts, and mitigation strategies*. MDPI. https://doi.org/10.3390/sci6010003

[4] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. https://www.propublica.org/article/machine-bias-risk-assessments-in criminal-sentencing