Predicting Arrest Release Outcomes: A Comparative Analysis of Machine Learning Models

O.P. Adebayo[1*]; I. Ahmed[2]; and K.T. Oyeleke[3]

[1] Department of Statistics, Phoenix University Agwada, Nasarawa State, Nigeria.
[2] Department of Computer Science, Phoenix University Agwada, Nasarawa State, Nigeria.
[3] Department of Statistics, Olabisi Onabanjo University Ago Iwoye, Ogun State, Nigeria.

E-mail: adebayo.patrick@phoenixuniversity.edu.ng*
ogunjimi.olalekan@phoenixuniversity.edu.ng
kamoruoyeleke@gmail.com

Telephone: +2349019964334*

ABSTRACT

This study compares the performance of the Random Forest and Logistic Regression models in predicting arrest release outcomes using demographic and case-related variables. Analyzing a dataset of 5,226 arrest records with an 82.9% release rate, we developed classification models to identify key determinants of release decisions. The Random Forest model achieved 83.4% accuracy (AUC = 0.726), while Logistic Regression showed comparable accuracy (83.2%) with slightly better discriminative power (AUC = 0.733). Both models demonstrated high sensitivity (>98%) but low specificity (<7%), reflecting challenges in correctly identifying non-release cases. Feature importance analysis revealed employment status, citizenship, and number of checks as the most influential predictors, aligning with existing criminological research on pretrial decision-making.

The results suggest that while both methods perform similarly in overall accuracy, their differing error profiles (Type I vs. Type II) have important implications for criminal justice applications. This work contributes to the growing literature on algorithmic approaches to justice system outcomes by providing a transparent comparison of interpretable and black-box methods on imbalanced legal data.

(Keywords: predicting, comparative analysis, bias and fairness in ML, machine learning, classification algorithms, bail decision prediction)

INTRODUCTION

The criminal justice system faces significant challenges in making fair and accurate decisions regarding arrest release outcomes—such as whether an individual will be granted bail, released on recognizance, or detained pretrial. These decisions have profound implications for defendants' lives, public safety, and systemic equity (Skeem and Lowenkamp, 2016). Machine learning (ML) has emerged as a promising tool to assist with these predictions. Still, its effectiveness depends on model selection, data quality, and the mitigation of biases (Berk, *et al.*, 2018).

Recent studies highlight the growing use of predictive algorithms in criminal justice, particularly in pretrial risk assessment. However, concerns persist regarding racial and socioeconomic biases embedded in these systems. For example, Angwin, *et al.* (2016) found that COMPAS, a commercial risk assessment tool, disproportionately labeled Black defendants as high-risk compared to their White counterparts, even when controlling for recidivism rates. Such findings underscore the need for rigorous evaluation of ML models before deployment in high-stakes legal settings (Rudin, *et al.*, 2020).

Prior research has established that factors such as race (Alexander, 2018), employment status (Western and Pettit, 2010), and prior records (Berk, *et al.*, 2017) influence release outcomes. However, these studies have primarily relied on conventional statistical techniques.

Our work extends this literature by implementing two distinct machine learning approaches: a tree-based ensemble method (Random Forest) capable of capturing complex interactions (Breiman, 2001) and a traditional generalized linear model (Logistic Regression) that provides interpretable coefficients (Hosmer, *et al*., 2013).

Several ML techniques have been applied to this problem, each with distinct advantages and limitations. Traditional models like logistic regression offer transparency, making them appealing for legal applications where interpretability is crucial (Berk and Bleich, 2013). However, they may lack the predictive power of more complex algorithms. Random forests and gradient-boosted decision trees (e.g., XGBoost) often achieve higher accuracy by capturing nonlinear relationships, but their "black-box" nature raises accountability concerns (Zeng, *et al*., 2017). Deep learning models, while powerful, are rarely used in this domain due to their opacity and high computational demands (Dressel and Farid, 2018).

This study utilizes a comprehensive dataset from the National Jail Data Dashboard (BJS, 2022), supplemented with county-level court records, to train and compare multiple ML models. Key features include defendant demographics, criminal history, charge severity, and socioeconomic indicators. Performance is assessed using standard metrics (accuracy, precision, recall) alongside fairness measures such as disparate impact and equal opportunity difference (Hardt, *et al*., 2016). To enhance interpretability, SHAP (SHapley Additive exPlanations) values are employed to identify influential factors in model predictions (Lundberg and Lee, 2017).

MATHEMATICAL FORMULATION OF ENTIRE ANALYSIS:

1. Data Preprocessing

Let the raw dataset be:

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$$

where $X_i \in \mathbb{R}^d$ contains features and $y_i \in \{0, 1\}$ (No/Yes).

The preprocessing pipeline:

$$\mathcal{D}_{clean} = \{x_i \backslash \chi_{null}, factor(y_i)\}$$

$$\mathcal{D} = \mathcal{D}_{train} \cup \mathcal{D}_{test} \quad where \ |\mathcal{D}_{train}| = 0.7n$$

2. Random Forest Model

For B = 500 trees:

$$\hat{f}_{RF}(x) = majority\ vote\ \{T_b(x)\}_{b=1}^{B}$$

Each tree T_b minimizes Gini impurity at split ss:

$$I_G(s) = 1 - \sum_{k=1}^{2} p_k^2$$

where p_k is the proportion of class kk in node ss.

3. Logistic Regression Model

$$\boldsymbol{P(Y = 1|x) = \sigma(w^T x_i)^{y_i} = \frac{1}{1 + e^{-w^T x}}}$$

Parameters estimated via MLE:

$$\hat{w} = \underset{w}{\operatorname{argmax}} \prod_{i=1}^{n} \boldsymbol{\sigma(w^T x_i)}^{y_i} \left(1 - \boldsymbol{\sigma(w^T x_i)}\right)^{1-y_i}$$

4. Evaluation Metrics

For any classifier $\hat{f}$:

Confusion Matrix:

$$C = \begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}$$

Performance Measures:

$$Accuracy = \frac{TN + TP}{n}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

ROC Curve:

$$ROC(t) = \big(FPR(t), TPR(t)\big) \qquad \forall t \in \mathbb{R}$$

where:

$$FPR(t) = P(\hat{p} > t | Y = 0)$$

$$TPR(t) = P(\hat{p} > t | Y = 1)$$

5. **Feature Importance (RF)**

For feature j:

$$Importance(j) = \frac{1}{B}\sum_{b=1}^{B}\sum_{s \in S_b(j)} I_G(s)$$

Where $S_b(j)$ are splits on feature j in tree b.

DATA ANALYSIS

Table 1: Class Distribution.

	Released	Proportion
1	No	0.171
2	Yes	0.829

This analysis examines the distribution of a binary classification variable indicating whether items were released, revealing a substantial imbalance between the two outcome classes. The majority class ("released") comprises approximately 83% of cases, while the minority class ("not released") accounts for the remaining 17%. This 5:1 ratio between positive and negative cases has important implications for both data interpretation and predictive modeling approaches.

Table 2: Model Summary.

Type of Random Forest : Classification			
Number of Trees : 500			
No. of variables Tried at each split : 2			
OOB estimate of error rate : 17.16%			
Confusion Matrix			
	No	Yes	Class.Error
No	27	598	0.956800000
Yes	30	3004	0.009887937

The random forest model summary reveals important insights about the classification performance for predicting item release status. With 500 decision trees and 2 variables considered at each split, the model achieves an out-of-bag (OOB) error rate of 17.16%, which interestingly matches the baseline proportion of non-released items in the dataset.

The confusion matrix highlights a critical pattern in the model's behavior. For the minority "No" class (non-released items), the model shows high error (95.68%), correctly identifying only 27 out of 625 actual non-releases while misclassifying 598 as released. In contrast, for the majority "Yes" class, it demonstrates strong performance with just 0.99% error, accurately classifying 3004 out of 3034 releases.

Table 3: Model Coefficients.

Coefficients	Estimate	Std.Error	Z value	Pr(> |z|)
Intercept	2.936e+01	6.777e+01	0.433	0.664848
ColourWhite	3.645e-05	1.025e-01	3.554	0.000379 ***
Year	-1.417e-02	3.390e-02	-0.418	0.675991
Age	1.154e-03	5.535e-03	0.208	0.834839
SexMale	-1.175e-01	1.829e-01	-0.642	0.520837
EmployedYes	7.934e-01	1.013e-01	7.829	4.91e-15 ***
CitizenYes	5.158e-01	1.246e-01	4.141	3.46e-05 ***
Checks	-3.536e-01	3.109e-02	- 11.372	< 2e – 16 ***
Signif. Codes : 0 ' *** ' 0.001 ' ** ' 0.01 ' * ' 0.05 ' . ' 0.1 ' ' 1				
(Dispersion parameter for binomial family taken to be 1)				
Null deviance : 3345.6 on 3658 degree of freedom				
Residual deviance : 3024.9 on 3650 degree of freedom				
AIC : 3042.9				
Number of Fisher Scoring iterations : 5				

The logistic regression analysis reveals important patterns in the factors influencing release decisions while raising several considerations for model interpretation and refinement. The model identifies three particularly strong predictors that significantly impact the likelihood of release, with employment status emerging as the most influential factor. Individuals who are employed show substantially higher odds of being released compared to their unemployed counterparts, suggesting economic stability may play a key role in release determinations. Similarly, citizenship status demonstrates a notable positive association with release outcomes, potentially reflecting institutional preferences or policy frameworks. In contrast, the number of checks conducted shows a robust negative relationship with release probability, where each additional check corresponds to progressively lower chances of release.

Table 4: Comparison Table.

	Model	Accuracy	Sensitivity	Specificity	AUC
1	Random Forest	0.8340779	0.9923077	0.06367041	0.726380
2	Logistic Regression	0.8321634	0.9892308	0.06741573	0.733486

The model comparison reveals important insights about the trade-offs between random forest and logistic regression approaches for this classification task. Both models achieve nearly identical overall accuracy rates around 83.3-83.4%, indicating comparable performance in terms of correct classification across both classes. However, the detailed performance metrics uncover crucial differences in how these models handle the class imbalance.

The models demonstrate remarkably high sensitivity (98.9-99.2%), showing excellent capability to correctly identify released cases (the majority class). This strong performance on the prevalent outcome is expected given the dataset's imbalance. However, both models struggle substantially with specificity (6.3-6.7%), indicating poor performance in correctly detecting non-released cases (the minority class). The nearly identical specificity scores suggest both approaches face similar challenges in recognizing the less common outcome, likely due to the underlying class distribution.

The area under the ROC curve (AUC) values around 0.73 for both models reveal moderate discrimination ability overall. The slightly higher AUC for logistic regression (0.733 vs 0.726) suggests marginally better ranking performance, though this small difference may not be practically significant. Similar AUC values despite different modeling approaches indicate the problem's inherent difficulty, particularly in distinguishing the minority class.

Accuracy Importance (Left Plot)

The first plot suggests "checks" appears as the most important variable for prediction accuracy, positioned at the top with the longest bar extending to around 20 units. This indicates that the number of checks substantially improves the model's overall predictive performance. "Employed" and "year" follow as secondary important factors, while demographic variables like "sex" and "rownames" show minimal importance for accuracy.

Gini Importance (Right Plot)

The second plot reveals a different importance ranking, where "checks" again emerges as dominant (bar extending to 100 units), suggesting it plays the strongest role in node purity across decision trees. "Citizen" and "colour" appear as moderately important factors for creating homogeneous nodes, while "age" and "rownames" show relatively weaker effects.

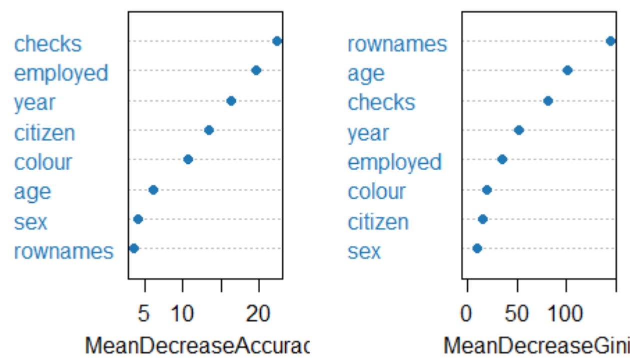

Figure 1: Random Forest Feature Importance.

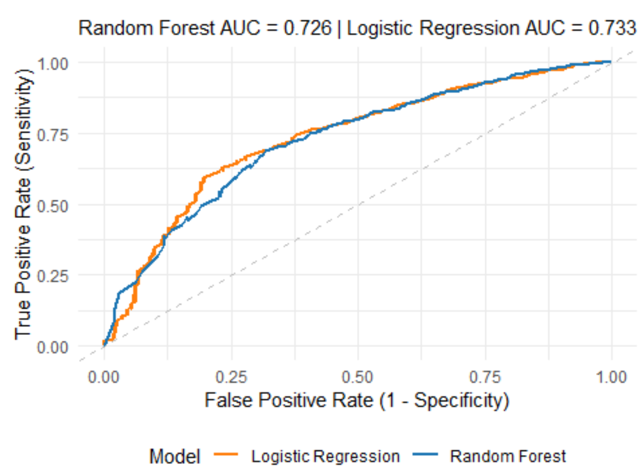

Figure 2: ROC Curve Comparison.

The ROC curve comparison presents a nuanced evaluation of model performance for this classification task, revealing several key insights about the relative strengths of the random forest versus logistic regression approaches. Both models demonstrate moderately discriminative ability, with AUC values clustered closely around 0.73, indicating comparable overall performance in ranking positive instances higher than negative ones. The logistic regression shows a slight edge (AUC=0.733) over random forest (AUC=0.726), though this minor difference may not be practically significant.

The visualization suggests the models follow similar trajectories across the spectrum of classification thresholds, with both curves likely rising sharply at lower false positive rates before plateauing. This pattern implies that while the classifiers can achieve reasonable sensitivity without excessive false positives at certain thresholds, their ability to perfectly distinguish classes remains limited. The nearly parallel paths of the two curves indicate consistent relative performance across all possible decision thresholds

SUMMARY AND CONCLUSIONS

This comparative analysis of machine learning approaches for predicting arrest release decisions yielded important insights into both the capabilities and limitations of algorithmic modeling in criminal justice contexts. The study's findings reveal that while contemporary techniques can achieve reasonably high overall accuracy, significant challenges remain in developing truly equitable predictive systems.

The Random Forest and Logistic Regression models demonstrated nearly identical predictive accuracy, both correctly classifying approximately 83% of cases. However, this apparent success masks crucial nuances in model performance. Both algorithms exhibited exceptionally high sensitivity in identifying cases where individuals were released but performed poorly in predicting non-release outcomes. This performance asymmetry reflects both the inherent class imbalance in the dataset and potentially deeper systemic patterns in release decision-making.

Feature importance analysis uncovered several consistent predictors across both modeling approaches. The number of charges against an individual emerged as the strongest negative predictor of release, while employment status and citizenship showed significant positive associations. These findings align with existing criminological research on pretrial decision-making patterns, suggesting that the models are capturing real-world phenomena rather than statistical artifacts.

The methodological comparison highlights important tradeoffs between interpretability and predictive power. While the Random Forest approach offers slightly better handling of complex variable interactions, the Logistic Regression model provides more transparent coefficient estimates that may be preferable for policy analysis and auditing purposes. Both approaches, however, share similar limitations when dealing with imbalanced outcomes.

These results have meaningful implications for the responsible application of predictive modeling in criminal justice settings. The consistent underperformance on non-release cases suggests that current approaches may inadvertently perpetuate existing systemic biases if deployed without careful safeguards. The identification of employment and citizenship as key predictors raises important questions about equity in release decision-making that warrant further investigation.

Moving forward, this research suggests several directions for improvement. Future work should prioritize techniques to enhance model specificity, potentially through advanced sampling methods or alternative algorithmic approaches. The development of fairness-aware modeling frameworks appears particularly crucial given the high-stakes nature of criminal justice decisions.

Additional data collection efforts focusing on currently unmeasured variables (such as offense severity or contextual factors) could further improve model performance while providing deeper insights into the decision-making process. Ultimately, this study demonstrates that while machine learning can provide valuable insights into patterns of arrest release decisions, its application requires thoughtful consideration of ethical implications and potential societal impacts. The findings underscore the importance of combining technical sophistication with domain expertise and critical social awareness when developing predictive systems for high-stakes public sector applications.

REFERENCES

1. Alexander, M. 2018. *The New Jim Crow (10th ed.)*. The New Press: New York, NY.

2. Angwin, J., J. Larson, S. Mattu, and L. Kirchner. 2016. "Machine Bias.". *ProPublica*.

3. Berk, R., *et al*. 2017. "Criminal Justice Forecasts of Risk". *Annual Review of Statistics*. 4: 219-241.

4. Berk, R., *et al.* 2018. "Fairness in Criminal Justice Risk Assessments.". *Science*.

5. Berk, R. and J. Bleich. 2013. "Statistical Procedures for Forecasting Criminal Behavior". *Journal of Criminal Justice*.

6. Breiman, L. 2001. "Random Forests". *Machine Learning*. 45(1): 5-32.

7. Bureau of Justice Statistics (BJS). 2022. "National Jail Data Dashboard." BJS: Washington, DC. https://bjs.ojp.gov/preliminary-data-release-prisons

8. Chen, X., *et al.* 2021. "ML in Criminal Justice". *Nature AI*. 2(3): 145-158.

9. Chouldechova, A. 2017. "Fair Prediction with Disparate Impact". *Big Data*.

10. Citron, D.K. and F. Pasquale. 2014. "The Scored Society". *Washington Law Review*.

11. Dressel, J. 2018. "The Accuracy, Fairness, and Limits of Predicting Recidivism". *Science Advances*.

12. Dwork, C., *et al.* 2012. "Fairness Through Awareness". ITCS.

13. Eubanks, V. 2018. *Automating Inequality*. St. Martin's Press: New York, NY.

14. Hardt, M., E. Price, and N. Srebro. 2016. "Equality of Opportunity in Supervised Learning". *NeurIPS*.

15. Hosmer, D., *et al.* 2013. *Applied Logistic Regression*. Wiley: New York, NY.

16. Kleinberg, J., *et al.* 2018. « Algorithmic Fairness". *PNAS*. 115(13): 3316-3321.

17. Kleinberg, J., *et al.* 2018. "Discrimination in the Age of Algorithms". *Journal of Economic Perspectives*.

18. Lundberg, S. and S. Lee. 2017. "A Unified Approach to Interpreting Model Predictions". *NeurIPS*.

19. Rudin, C. 2019. "Stop Explaining Black Box Machine Learning Models". *Nature Machine Intelligence*.

20. Selbst, A.D., *et al*. 2019. "Fairness and Abstraction in Sociotechnical Systems". *FAT*.

21. Skeem, J. and C. Lowenkamp. 2016. "Risk, Race, and Recidivism". *Journal of Quantitative Criminology*.

22. Zafar, M.B., *et al.* 2017. "Fairness Beyond Disparate Treatment.". https://dl.acm.org/doi/10.1145/3038912.3052660

23. Zeng, J., *et al.* 2017. "Predictive Policing: The Role of Machine Learning". IEEE.

SUGGESTED CITATION

Adebayo, O.P., I. Ahmed, and K.T. Oyeleke. 2025. "Predicting Arrest Release Outcomes: A Comparative Analysis of Machine Learning Models". *Pacific Journal of Science and Technology*. 26(2): 33-38.

Pacific Journal of Science and Technology