VOL. 13 NO. 9,
AUGUST, 2026
E-ISSN 3026-958X
P-ISSN 3027-1169

African Scholar Publications & Research International
www.africanscholarpub.com

Journal of Science Innovation & Technology Research (JSITR)

Fairness and Bias Detection in Artificial Intelligence-Driven Decision-Making in Employment Processes using Machine Learning Model

***Francisca Nonyelum Ogwueleka; **Uchenna Igboeli; & ***Isa Ibrahim Wadda**

*University of Abuja. francisca.ogwueleka@uniabuja.edu.ng. **University of Abuja. uchenna.igboeli@uniabuja.edu.ng ***University of Abuja. ibrahimisa959@gmail.com

DOI: https://doi.org/10.70382/ajsitr.v13i9.081

Abstract

This study investigated fairness and bias detection in Artificial Intelligence (AI)-driven employment decision-making using machine learning models. The research aimed to determine whether AI-based recruitment systems make fair hiring decisions by evaluating key fairness. Additionally, it conducted an in-depth study on the application of AI in hiring, focusing on how it might save hiring expenses by automating the screening of numerous applicants. A machine learning-based model was developed and implemented using Python to assess fairness and bias in AI recruitment processes. The study utilized a dataset of 1,000 applicant records containing demographic and professional attributes such as gender, education, skills, and years of experience. Random Forest (RForest), Gradient Boosting Machine (GBoosting), and Support Vector Machine (SVector) algorithms were evaluated based on predictive performance and fairness. The outcome of the work shows that the SVector achieved the highest classification accuracy of 87.0% and demonstrated the best overall fairness performance across all evaluated metrics. The SVector model exhibited lower levels of bias relating to gender, location, experience, and disability inclusion. Feature importance analysis showed that years of experience was the most influential factor in hiring decisions across all models, while the influence of gender varied between models. This work examined the importance of balancing predictive accuracy with fairness when deploying AI systems in recruitment. Based on the results, the SVector model was identified as the most suitable approach due to its strong performance and reduced bias. To further improve fairness in AI recruitment systems, the study recommends continuous monitoring and regular fairness audits. These measures can help ensure more ethical, transparent, and equitable AI-driven hiring practices.

Keywords: Artificial Intelligence, Fairness, Bias, Support Vector, Random Forest, Gradient Boosting

Introduction

Taking decisions based on AI algorithms can improve efficiency, accuracy, and objectivity by analyzing large volume of data and spotting patterns that humans might miss. AI integration into decision-making processes has become more widespread in recent years across a variety of sectors, from criminal justice to finance and healthcare. But the broad use of these algorithms also brings up important societal issues that need careful thought and research. In order to increase efficiency and objectivity, AI-driven algorithms are frequently employed to automate processes (Lepri et al., 2017).

If suitable mitigation techniques are not used, artificial intelligence systems may learn and reinforce historical biases included in training datasets, leading to unjust or discriminating choices (Eubanks, 2018). Furthermore, many AI models lack transparency, which makes it challenging for stakeholders to comprehend decision-making processes and establish accountability for them (Sandvig et al., 2016).

These algorithms' possible biases and moral dilemmas have drawn criticism (Cath et al., 2017). Therefore, strong governance frameworks are necessary for appropriate adoption of AI in legal decision-making through efficient regulation and supervision (Raso et al., 2018).

AI recruiting algorithms, like the Application Tracking System (ATS), have broad societal ramifications that can profoundly affect both individuals and communities. For example, the application of AI in Law has been linked to racial inequities in sentence and parole decisions (Angwin et al., 2016). The primary guidelines for automated decision-making systems are shown in Figure 1., In a similar vein, unequal access to medical services and treatment delivery may occur from the application of AI in healthcare (Char et al., 2018).



Figure 1. Automated Decision-making Systems (Angwin et al., 2016)

Decision-making procedures in a number of industries, including criminal justice, finance, and employment, are increasingly utilizing artificial intelligence (AI). (Cath et al., 2018). In order to give policymakers and business professionals practical insights and suggestions for improving fairness and equity in AI-driven employment practices, this study investigates the detection of bias and fairness in AI-driven systems.

Statement of Problem

Using technologies to automate the screening of job applications, such as Applicant Tracking Systems (ATS) and Machine Learning (ML) algorithms. These tools help businesses reduce hiring costs and boost productivity by quickly processing enormous volumes of resumes. However, concerns about fairness, accountability, transparency, and bias in AI-generated hiring decisions have emerged.

Prior research on AI hiring has mostly focused on surveys and statistical analysis, which has limited understanding of how machine learning algorithms actually behave. Furthermore, little empirical study has been done on algorithmic bias and fairness in hiring processes.

To bridge this gap, this study developed a machine learning framework to evaluate bias and fairness in AI-based hiring practices. In the study, the algorithms were compared. Their effectiveness was assessed using fairness metrics such treatment equality, equal opportunity, demographic parity, and predictive parity.

The research provides a dataset-based comparative analysis of algorithmic fairness and bias, offering deeper insights into AI recruitment decisions. The goal was to identify the most suitable algorithm for hiring processes while balancing predictive accuracy with fairness.

Objectives of the study

In order to give policymakers and business professionals practical insights and suggestions for improving fairness and equity in AI-driven employment practices, the study's goal is to investigate Fairness and Bias Detection in AI-Driven Decision Making in Employment Processes Using Machine Learning Model.

The objectives are to:

i. Using the specified ML classifiers, develop a model that examines potential biases and moral dilemmas associated with the application of AI-driven decision-making in the employment process.
ii. Evaluate the models' fairness and bias using standards fairness metrics: equal opportunity, treatment equity, predictive parity, and demographic parity.

Related Works

The role AI in boosting productivity and facilitate decision-making in hiring procedures has drawn a lot of interest. However, it is challenging to comprehend how judgments are produced and if they are impartial or biased. This study suggests a paradigm for employing machine learning models to examine bias and fairness in AI-driven hiring. It looks at important aspects of AI including governance, which affect public confidence and give rise to ethical and social inequity concerns.

The conceptual framework is illustrated in Figure 3.

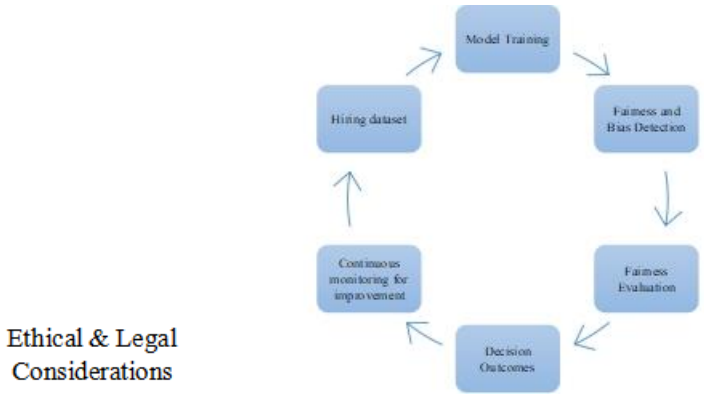


Figure 3. Conceptual framework diagrammatic illustration (Sandvig et al., 2016).

Automated Decision-Making Systems (ADMS) are systems that leverage data, machines, and algorithms to make decisions across diverse sectors, including public administration, business,

healthcare, and employment. As shown in Figure 3, these systems function with increasing degrees such as seeing, orienting, making decisions, and acting. They can also infer predictions, content, suggestions, or decisions from input data to affect real-world or virtual environments. The widespread adoption of ADMS is transforming operations by offering substantial gains in efficiency, speed, accuracy, and consistency through the processing of vast datasets and identification of complex patterns beyond human cognitive capabilities (Marabelli *et al.*, 2021).

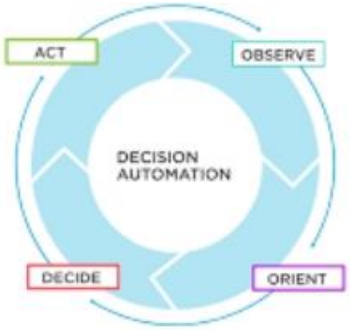


Figure 2. Decision Automation (Marabelli *et al.*, 2021)

All AI systems are fundamentally machine-based, depending on computational operations including both hardware and software components. The infrastructure required for data processing, model training, predictive modeling, and extensive automated decision-making is provided by this synergy. The main distinction between AI agents and other AI systems is their operational capabilities and architectural design (Gonzalez et al., 2019). AI agents are more autonomous in pursuing their objectives than conventional systems, which follow explicit instructions and predetermined parameters.

A major conceptual shift may be seen in the transition from simple automation to increasingly autonomous intelligence. Modern AI systems have the ability to see, reason, and act on their own, frequently modifying their behavior in response to feedback from their surroundings. In contrast, early automated systems mainly followed predetermined rules. This evolution signifies a move towards greater sophistication and independence in decision-making processes (Marabelli et al., 2021).

Rule-based Systems

Rule-based systems are a fundamental type of AI-driven decision-making algorithm that bases its decisions on a predefined set of rules. These systems are widely used in many different areas, including manufacturing, healthcare, and finance, because to their interpretability and ease of use (Ligęza & Nalepa, 2011). Figure 4, shows the flow of a rule-based system execution process.

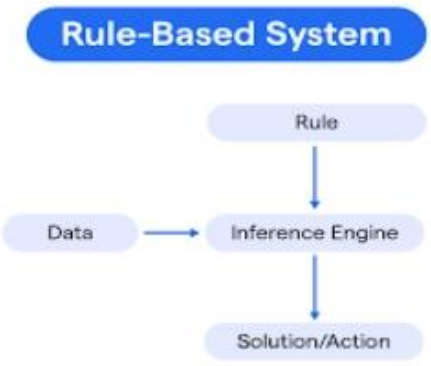


Figure 4. Rule Based System (Miller, 2023)

Fairness through Unawareness

If the decision-making process does not specifically take into account protected traits, the decision-maker is said to accomplish fairness through unawareness (Chen et al., 2019). If the hiring decision-maker in our case disregards a candidate's gender, this can be easily accomplished by merely concealing this characteristic during the training phase. Because of the interconnectedness with other factors, one may be discriminating even if they are unaware of the protected trait, making this method naïve.

Preexisting discrimination and inequality may be maintained or even made worse by biased algorithms. Similar to this, a lending algorithm that is biased against some groups may make it harder for those

groups to obtain credit, which would exacerbate already-existing economic inequality. Machine discrimination has the potential to create new types of prejudice in addition to sustaining preexisting ones.

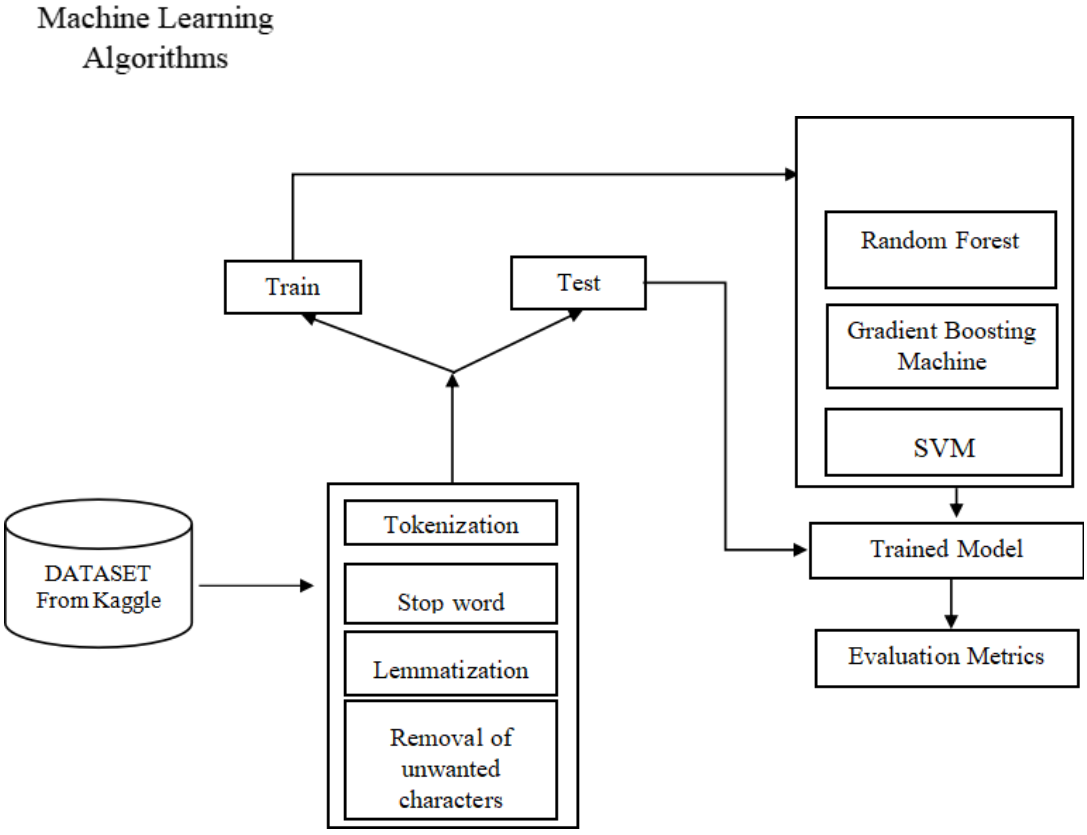


Figure 5. System Design

METHODOLOGY

The first step in the research process is data collection. Since textual data is usually unstructured, many common pre-processing techniques must be used to migrate that data into a system that can be given into the classification system for pre-processing. Using the widely used and standard feature extraction technique of TF-IDF. The cleaned dataset is then reduced to the specific features that describe the recruiter's decision to hire and reject. The data was then divided into 80:20 ratios for testing and training, respectively, and fed into the models for assessment.

Data Collection

A public (hiring dataset) with demographic features such; name, gender, age, location, skills, year of experience, certification, job role and recruiter's decision, salary expectation and completed project count was used for the research obtained from Kaggle. The gender was considered the protected attribute.

The dataset has a record of 1000 entries with features considered such as;

i. Resume_ID: Unique identifier
ii. Demographic Status: Such as candidate's name, gender, age, location and disability status.

iii. Skills: List of relevant technical skills
iv. Years of Experience: Total work experience in years
v. Education: Highest educational qualification
vi. Certifications: Relevant industry certifications
vii. Job Role: Target job position
viii. Projects Count: Number of projects completed
ix. AI Score (0-100): AI-based resume ranking score
x. Recruiter Decision: Hire or Reject

Data Pre-processing

Data processing is a form of noise reduction technique normally applied to collected data before subjecting it to any form of analysis.

Online data like employee data, status and other related features have some form of open-endedness and more prone to noise (Han et al., 2012). The data was created in a clean format that can improve the ensemble models and guarantee that their performance is not affected by the existence of noisy and unbalanced data in order to obtain good performance from the models. The pre-processing techniques as listed in Figure 5, are further discussed as follows;

Feature Extraction

Textual evaluations typically offer a wealth of insightful information that might assist readers in making wise choices. However, the implementation of feature extraction methods is significant in terms of vectorizing the phrases. These are the essential prerequisites for vectorizing a sentence:

i. Have a sparse matrix, i.e. one containing a lot of 0's, which signifies that the word did not occur at all in the review sentences.
ii. Retrain the linguistic information present in the text review when converted into a vectors.
iii. Term Frequency

$$TF = \frac{Number\ of\ times\ the\ term\ appears\ in\ the\ document}{Total\ number\ of\ terms\ in\ the\ document}$$

iv. Inverse Document Frequency

$$IDF = \log\frac{Total\ number\ of\ documents}{Number\ of\ documents\ containing\ the\ term}$$

Model Performance Evaluation Metrics

i. Accuracy:

$$Accuracy = \frac{correct\ classifications}{total\ classification} = \frac{TP + TN}{TP+TN+FP+FN} \quad (3.1)$$

ii. Precision:

$$\text{Precision can be expressed mathematically as} = \frac{TP}{TP + FP} \quad (3.2)$$

iii. Recall:

$$Recall = \frac{TP}{TP+FN} \quad (3.3)$$

iv. F-measure:

F-measure = $\frac{2 \times (precision \times recall)}{precision + recall}$ (3.4)

Fairness Performance Evaluation Metrics

Aside the overall performance metrics there are also fairness and bias metrics used in the research as discussed below;

Demographic Parity:

$$P(\hat{Y}=1|A=0) = P(\hat{Y}=1|A=1)$$

(3.5)

Equal Opportunity:

$TPR_{protected} = TPR_{non\text{-}protected}$ i.e $\frac{TP}{TP+FN}\Big|_{protected} = \frac{TP}{TP+FN}\Big|_{non\text{-}protected}$

(3.6)

$Precision_{protected} = Precision_{non\text{-}protected}$ i.e $\frac{TP}{TP+FN}\Big|_{protected} = \frac{TP}{TP+FN}\Big|_{non\text{-}protected}$ (3.7)

Treatment EqualityFormula:

$$\frac{P(\hat{Y}=0|Y=1,\ A=a)}{P(\hat{Y}=1|Y=0,\ A=a)} = \frac{P(\hat{Y}=0|Y=1,\ A=b)}{P(\hat{Y}=1|Y=0,\ A=b)}$$

(3.8)

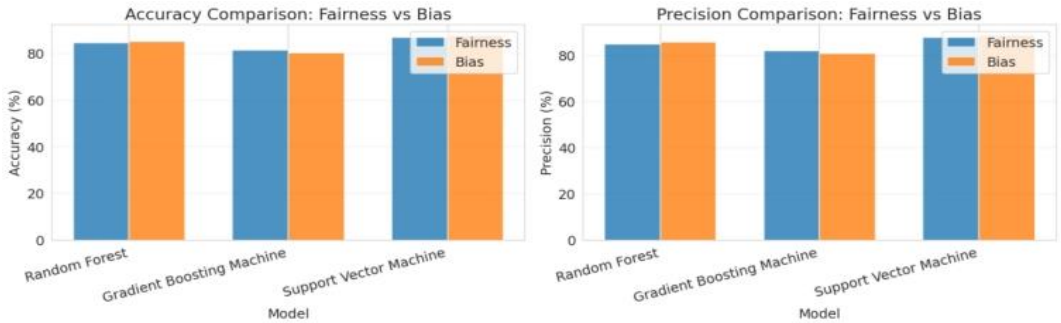
Results and Discussion

Implementation encapsulates all the processes in transforming the idea presented in the abstract into the completed work. It gives detailed implementation and evaluation results of the ensemble classifiers of GBoosting, RForest and SVector for the fairness and bias detection on the public hiring datasets adopted and provide the evaluation of the system's performance using the performance metrics defined in previous chapters.

Table 1. Models Performance Table

Model	**Accuracy (%)**		**Precision (%)**		**Recall (%)**		**F1 Score (%)**	
	Fairness	Bias	Fairness	Bias	Fairness	Bias	Fairness	Bias
RForest	84.50	85.05	84.99	85.74	84.50	85.05	84.72	85.36
GBoosting	81.50	80.37	82.07	80.81	81.50	80.37	81.76	80.58
SVector	87.00	87.85	87.90	88.76	87.00	87.85	87.35	88.21

Source: AI Resume Screening dataset from kaggle



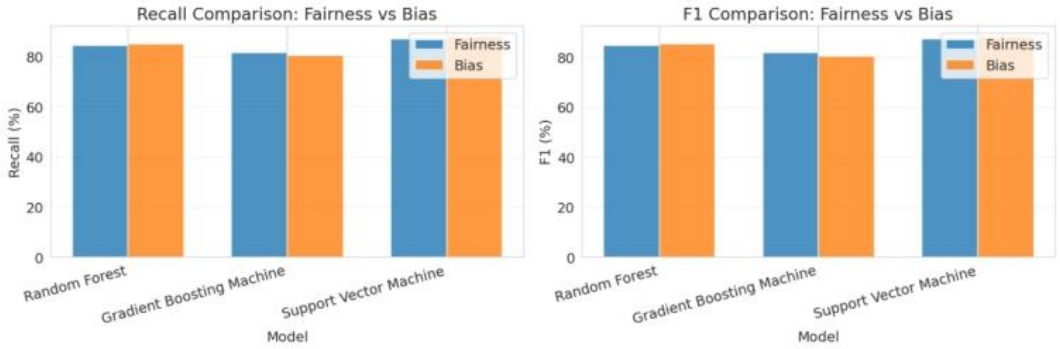


Figure 6. Chart on Model Performance summary

The comparative analysis indicates clear differences in predictive capability among the three classifiers. SVector consistently achieved superior values across all metrics, suggesting a stronger ability to distinguish between fair and biased hiring outcomes. RForest maintained stable performance, whereas GBoosting produced comparatively lower predictive effectiveness.

With accuracy between 87.00% and 87.85%, precision between 87.90% and 88.76%, recall between 87.00% and 87.85%, and F1 scores between 87.35% and 88.21%, SVector outperformed all other models. These findings show that SVector is the best model for balanced classification in employment-related AI systems since it showed excellent robustness and reliability in identifying both fair and biased decisions.

The RForest model also exhibited strong and stable performance, with accuracy between 84.50% and 85.05%, precision from 84.99% to 85.74%, recall between 84.50% and 85.05%, and F1 scores ranging from 84.72% to 85.36%. This consistency suggests that RForest can effectively handle fairness and bias detection with minimal performance fluctuation. In contrast, the GBoosting showed slightly lower results across all metrics, with accuracy between 80.37% and 81.50% and F1 scores from 80.58% to 81.76%, indicating relatively weaker predictive strength. Overall, while all three models performed satisfactorily, SVector outperformed the others, followed by RForest, with GBoosting ranking last but still maintaining acceptable performance for bias and fairness assessment.

Table 2. Feature Importance

Model	**Feature Importance Fairness**			**Feature Importance Bias**		
Feature	GBoosting	RForest	SVector	GBoosting	RForest	SVector
Age	0.167061	0.165106	0.000124	0.193636	0.180941	0.000141
Certifications	0.027757	0.029672	0.367271	0.026416	0.025339	0.318803
Education	0.059506	0.075830	0.062954	0.049526	0.077676	0.069014
Years of experience	0.544527	0.508836	0.726136	0.444834	0.438328	0.862726
Gender	0.024699	0.031507	0.353167	0.021842	0.034769	0.375282
Job Role	0.074035	0.069080	0.001974	0.063534	0.059147	0.002146
Location	0.061797	0.072988	0.056398	0.066506	0.072949	0.062421
Skills Count	0.040618	0.046982	0.239951	0.041548	0.049843	0.252077

Source: AI Resume Screening dataset from kaggle

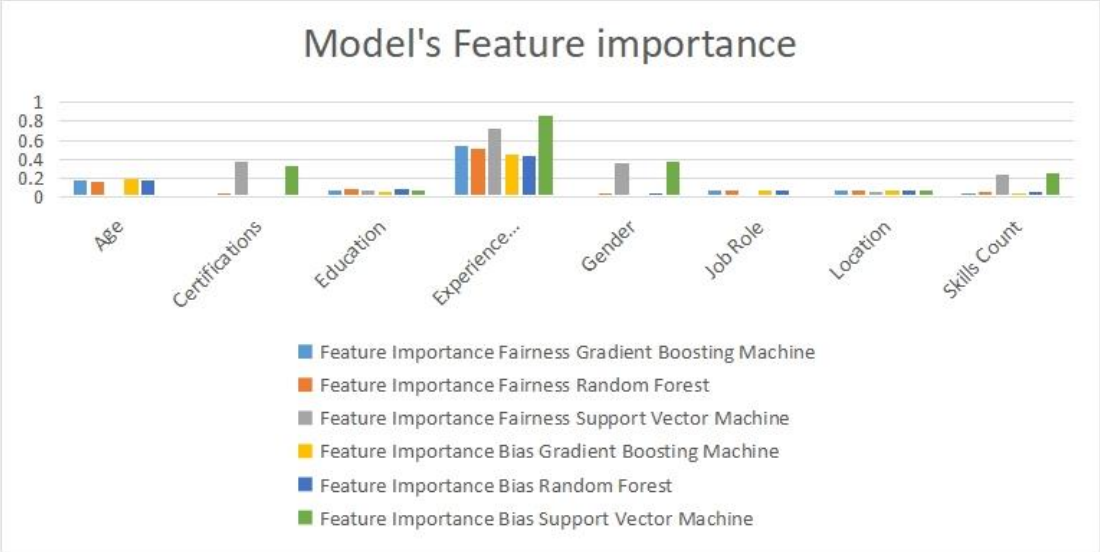


Figure 7. Chart of the Model's feature performance comparison

Feature importance analysis demonstrates that professional experience contributes more to model predictions than any other attribute evaluated. Across the three classifiers, this variable consistently received the highest importance scores, indicating that employment decisions were driven primarily by accumulated work experience rather than demographic characteristics. Years of experience consistently shows the highest importance fairness scores (0.54, 0.51 and 0.73) for GBoosting, RForest and SVector respectively under feature importance fairness. This suggests that the number of years of experience is a major factor in the models' hiring decisions. Skills Count and Certifications (0.37) also appear to be important features, particularly for the SVector, this indicates that the number of skills and certifications listed by a candidate influences the model's prediction. While importance of other features such age, education, Job Role, and Location varies across the models. For instance, job role has displaced as more important for GBoosting (0.07) and RForest (0.07) than for SVector (0.00) whereas gender shows some level of importance, particularly for the SVector (0.4). This is a critical finding when analyzing for bias, as it suggests the model will be using gender information in its predictions. Ultimately, this feature importance analysis highlights that years of experience is the most influential factor in the models' hiring predictions, with skills count and certifications count also playing significant roles

Table 3. Fairness Performance Metrics

Model	Demographic Parity	Equal Opportunity	Predictive Parity	Treatment Equality
RForest	0.038891	0.038889	0.069048	0.142857
GBoosting	0.058989	0.205556	0.169856	0.185714
SVector	0.051050	0.027778	0.033126	0.000013

Source: AI Resume Screening dataset from Kaggle

Examination of the fairness indicators reveals that each classifier behaves differently across protected groups. Although RForest maintained relatively balanced values, SVector produced the smallest

disparities for Equal Opportunity, Predictive Parity, and Treatment Equality, indicating greater consistency in treating applicants irrespective of protected characteristics

RForest has moderate values across all metrics demographic parity (0.04), equal opportunity (0.06), predictive parity (0.07), and treatment equality (0.14) indicating consistent but cautious fairness sensitivity. The GBoosting showed slightly higher disparities, particularly in equal opportunity (0.21) and predictive parity (0.17), suggesting that while it captures fairness differences more strongly, it also introduce variability across demographic groups. Conversely, the SVector displayed the lowest fairness variation overall, with minimal scores in equal opportunity (0.03), predictive parity (0.03), and treatment equality (0.00), implying that it enforces stricter uniformity in decisions but may underrepresent nuanced fairness considerations. Overall, RForest achieved the most balanced fairness distribution, GBoosting showed higher fairness sensitivity, and SVector leaned toward consistency with limited fairness differentiation. Ultimately, in summary Based on these fairness metrics, the SVector appears to be the fairest model in terms of treating group protection. The graphical illustration is presented in figure 8.

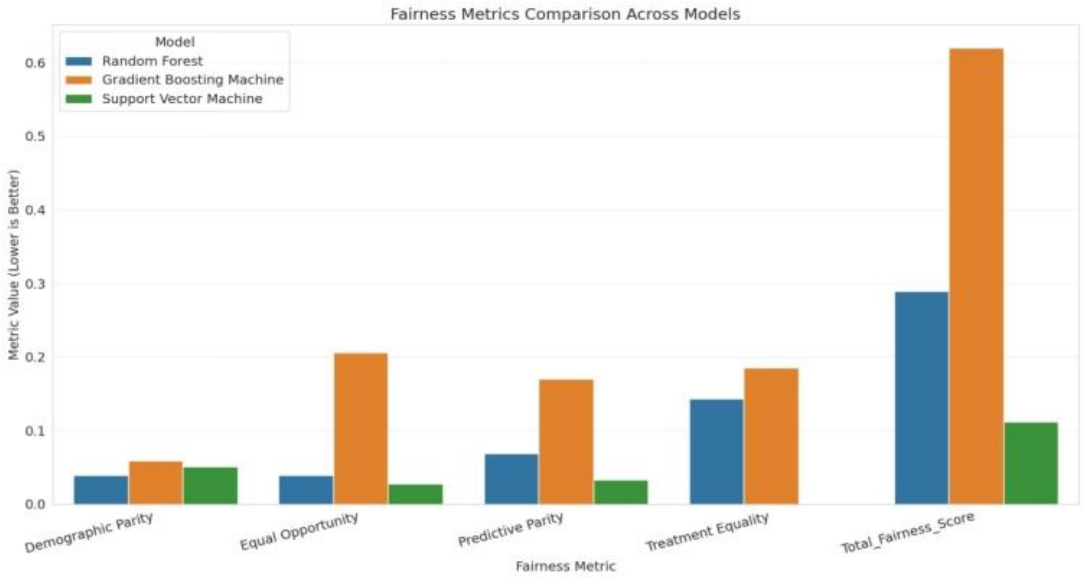


Figure 8. Chart on Fairness Metrics Comparison

Table 4. Bias Detection Metrics performance

Model	Demographic Parity	Equal Opportunity	Predictive Parity	Treatment Equality
RForest	0.172230	0.038889	0.069048	0.142857
GBoosting	0.249363	0.205556	0.169856	0.185714
SVector	0.206420	0.027778	0.033126	0.000013

Source: AI Resume Screening dataset from Kaggle

In evaluating bias across the models, Table 4.8 shows the GBoosting exhibited the highest bias levels overall, with notable values in demographic parity (0.25), equal opportunity (0.21), predictive parity (0.17), and treatment equality (0.19), indicating a stronger tendency toward unequal outcomes among demographic groups. The RForest model showed moderate bias levels, particularly in demographic

parity (0.17) and treatment equality (0.14), while maintaining lower values for equal opportunity (0.04) and predictive parity (0.07), reflecting relatively controlled bias with occasional deviations. In contrast, the SVector demonstrated the lowest bias levels across all metrics, especially with treatment equality at 0.00 and minimal values for equal opportunity and predictive parity (0.03 each), suggesting higher consistency and fairness in decision outcomes. Overall, SVector emerged as the least biased model, RForest maintained balanced bias control, and GBoosting displayed the most pronounced bias tendencies among the three classifiers. Therefore, the bias metrics confirm that the SVector model demonstrates the least amount of bias with respect to the protected group (female). This aligns with the findings from the fairness metrics analysis. The graphical illustration of the data in the Table 4. is illustrated in figure 9.

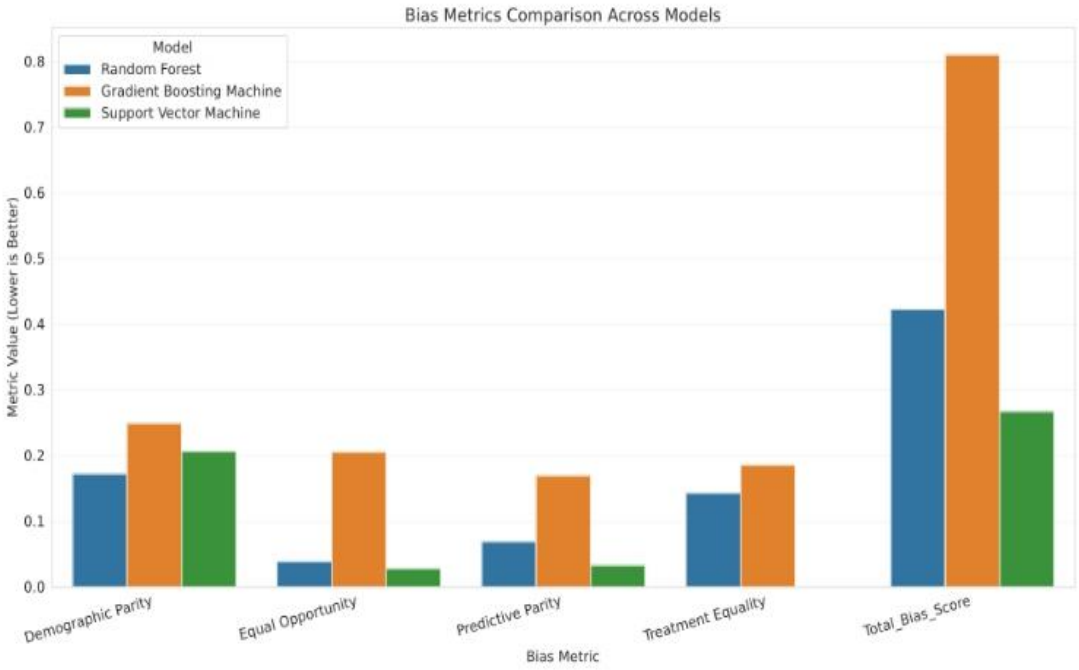


Figure 9. Chart on Bias Metrics Comparison

Comparison to the Previous Studies

Studies like Gonzalez *et al.*, (2019), Jain *et al.*, (2024), Bankins & Formosa (2023), and Min (2023) used statistical tools such as SPSS and correlations and for the justification of their studies to evaluate the overall nature of bias in the systems while this studies employed the use of technical tool - python programming - for the implementation and machine learning algorithms. This enables us to make justification using the trending model performance evaluation metrics to determine the demographic parity, equal opportunity, predictive parity, and treatment equality using a public hiring (AI resume screening) dataset obtained from kaggle repository online to provide an experimental dataset based results.

Conclusion and Recommendations

The three machine learning classifiers' prediction capacities differ noticeably. The SVector consistently performed the best, demonstrating its greater capacity to correctly categorize recruitment outcomes.

While the GBoosting gave relatively lower performance values across the assessment metrics, the RForest model also offered steady and dependable predictions. These results are in line with the suggestion that machine learning models should be evaluated using a variety of performance metrics rather than only classification accuracy (Han et al., 2012).

With precision values ranging from 87.90% to 88.76% and prediction accuracy ranging from 87.00% to 87.85%, SVector demonstrated the strongest generalization ability among the assessed models. GBoosting yielded respectable but somewhat poor predicting results, while RForest consistently performed well across all criteria. Overall, the results show that, among the datasets utilized in this study, SVector offers the most dependable classification performance for AI-driven employment decision-making.

The fairness evaluation demonstrates that each classifier produced different levels of disparity across the protected groups. Although RForest maintained relatively balanced fairness values, the SVector consistently recorded the smallest disparities for Equal Opportunity, Predictive Parity, and Treatment Equality. Lower disparity values indicate that the classifier produced more consistent outcomes across demographic groups, suggesting a lower likelihood of discriminatory treatment (Lepri et al., 2018).

RForest displayed moderate fairness across all four evaluation metrics, indicating a stable but conservative approach to decision-making. In comparison, GBoosting recorded larger disparities, particularly for Equal Opportunity and Predictive Parity, implying greater variation in prediction outcomes between protected groups. Conversely, the SVector achieved the lowest disparity values overall, reflecting greater consistency in treating applicants with different demographic characteristics. These findings support the view that fairness should be evaluated using multiple complementary metrics rather than relying on a single indicator (Chen et al., 2019).

The bias assessment further reinforces the findings obtained from the fairness analysis. GBoosting exhibited the highest bias values across most evaluation metrics, indicating a greater tendency to produce unequal outcomes among demographic groups. RForest demonstrated comparatively lower bias and maintained relatively balanced performance across the protected attributes evaluated.

SVector consistently recorded the lowest bias values for Equal Opportunity, Predictive Parity, and Treatment Equality, indicating that its predictions were less influenced by protected characteristics. These findings suggest that SVector achieved the best balance between predictive performance and equitable treatment of applicants. Previous studies have shown that improving accountability, transparency in modern decision-making systems requires eliminating algorithmic bias. (Lepri et al., 2018; Raso et al., 2018). Consequently, SVector seems to be the best algorithm among the assessed classifiers for assisting equitable AI-driven hiring selections.

Recommendations

The following recommendations are hereby proposed:

i. Organizations using AI-based hiring systems should prioritize models like SVector, which demonstrate minimal bias and high fairness consistency.
ii. Regular fairness audits should be conducted to monitor and mitigate model bias in employment decision-making processes.
iii. Datasets used in AI recruitment systems should be continuously reviewed and balanced to prevent historical or structural bias.
iv. Disability status identity should be captured in the datasets so as to reflect the inclusion criteria as advocated at the world convention addressing disabilities.

REFERENCES

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bankins, S., & Formosa, P. (2023). The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work. *J Bus Ethics* 185, 725–740. https://doi.org/10.1007/s10551-023-05339-7

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2017). *Artificial intelligence and the "good society": The US, EU, and UK approach*. *Science and Engineering Ethics*, *24*(2), 505–528. https://doi.org/10.1007/s11948-017-9901-7

Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care: Addressing ethical challenges. New England Journal of Medicine, 378(11), 981–983. https://doi.org/10.1056/NEJMp1714229

Chen, J., Kallus, N., Mao, X., Svacha, G., & Udell, M. (2019). *Fairness under unawareness: Assessing disparity when protected class is unobserved*. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 339–348). ACM. https://doi.org/10.1145/3287560.3287594

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Gonzalez, M. F., Capman, J. F., Oswald, F. L., Theys, E. R., & Tomczak, D. L. (2019). "Where's the IO?" Artificial intelligence and machine learning in talent management systems. Personnel Assessment and Decisions, *5*(3). https://doi.org/10.25035/pad.2019.03.005

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Jain, R., Patidar, N., Mishra, S., Prajapati, D., Solanki, A., Suthar, R., Patel, K., & Patel, H. (2024). Transparency in AI decision making: A survey of explainable AI methods and applications. *Advances in Robotics and Technology*, *8*(1), 21. https://doi.org/10.23880/art-16000110

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. Philosophy & Technology, 31(4), 611-627. https://doi.org/10.1007/s13347-017-0279-x

Marabelli, M., Newell, S., & Handunge, V. (2021). The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *The Journal of Strategic Information Systems*, *30*(1), Article 101652. https://doi.org/10.1016/j.jsis.2021.101652

Miller, J. (2023). Ease of implementation and maintenance in rule-based systems. *Springer*, *77*(1), 71–88.

Ligęza A., & Nalepa G.J., (2011). A study of methodological issues in design and development of rule-based systems: proposal of a new approach. WIREs Data Mining and Knowledge Discovery, Volume1, Issue2. Pages, 117-137

Min, A. (2023). Artificial intelligence and bias: Challenges, implications, and remedies. *Journal of Social Research*, *2*(11), 3808–3817. DOI:10.55324/josr.v2i11.1477

Raso, F. A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018). Artificial intelligence & human rights: Opportunities & risks. Berkman Klein Center for Internet & Society Research Publication, http://nrs.harvard.edu/urn

Sandvig, C., Hamilton, K., Karahalios, K. & Langbort, C. (2016). When the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software. International Journal of Communication, 10, 4972–4990