Final Report: Studying Trade-off between Accuracy and Fairness in AI-based Employee Performance Evaluation

[redacted] [redacted] [redacted]
[redacted] [redacted] [redacted]
[redacted]

1 Abstract

This study investigates the trade-off between accuracy and fairness in AI-based employee performance evaluation. Utilizing a dataset from INX Future Inc., comprising performance records of 1200 employees, we scrutinize biases and their impact on evaluation outcomes. Through a methodology encompassing data preprocessing, exploratory data analysis, clustering, and model development using Random Forest and CatBoost algorithms, we aim to maximize accuracy and enhance fairness in performance prediction. Our findings emphasize the intricate relationship between algorithmic interventions and fairness metrics, pointing out the significance of tailored model development for achieving equitable results. This research adds valuable insights to discussions on fairness in AI-driven employee evaluation systems and points toward potential areas for further investigation.

2 Introduction

Companies are increasingly turning to AI to evaluate employee performance, leveraging sophisticated machine learning models to analyze and interpret complex datasets. While these AI systems hold the promise of objectivity and efficiency, they also pose significant challenges, particularly in terms of fairness and bias. In this study, we plan to investigate the trade-off between model's accuracy and fairness in AI-based employee performance evaluation. By prioritizing both accuracy and fairness, we can harness the power of AI to improve efficiency and objectivity in evaluations while maintaining ethical standards and promoting workplace equality.

We are using the dataset from INX Future Inc., which consists of performance appraisal records for 1200 employees featuring 28 parameters such as education level, environment satisfaction, job involvement, job satisfaction, performance rating, relationship satisfaction, and work-life balance.

Mitigating bias in the dataset is vital for ensuring the model's accuracy. In the dataset there are three main forms of bias—functional bias (societal environment-caused bias), age bias, and gender bias that are of particular concern. Firstly, functional Bias stems from societal factors influencing employee performance evaluation. Features such as educational level, marital status, department, educational background, and distance from home may introduce bias due to societal norms and disparities. Secondly, prejudices or stereotypes based on age reflected in the "Age" feature may lead to biased evaluations of older or younger employees regarding their capabilities and potential for growth, called Age Bias. Lastly, differential treatment based on gender identity, indicated by the "Gender" feature, can result in biased evaluations affecting opportunities and advancement for male and female employees, called Gender Bias.

[redacted]

3 Related Works

The integration of AI in Human Resource Management (HRM) signifies a transformative shift across its six dimensions: strategic planning, recruitment, training and development, performance management, compensation management, and human relations management (Noe et al., 2006). This fusion promises enhanced efficiency in HRM tasks, from strategic alignment and talent acquisition to employee evaluation.

In recent studies, traditional machine learning techniques have been used in the performance evaluation field. In one of the studies, surveys were conducted to collect various details from individuals, including personal, job-specific, and socioeconomic information like gender, work experience, managerial level, income, and educational background. This data was then used to classify employee performance into three categories: low, medium, and high. This classification was done using different machine learning techniques such as Naive Bayes, Random Forest, Support Vector Machine, Multi-layer Perceptron, and Logistic Regression to evaluate and compare the effectiveness of these methods in assessing employee performance (A. Lather et al., 2019). Another research attempted to use the XGBoost ensemble learning model on a collection of employee datasets from an multinational corporations from the Chennai database (Sujatha and Dhivya, 2022).

Despite numerous studies employing machine learning for evaluating employees, the intersection of bias with these techniques has not been extensively explored. In the above papers, the concept of bias was not addressed, reflecting that algorithm and bias were likely to be studied separately. However, the fact that these 2 concepts go hand in hand

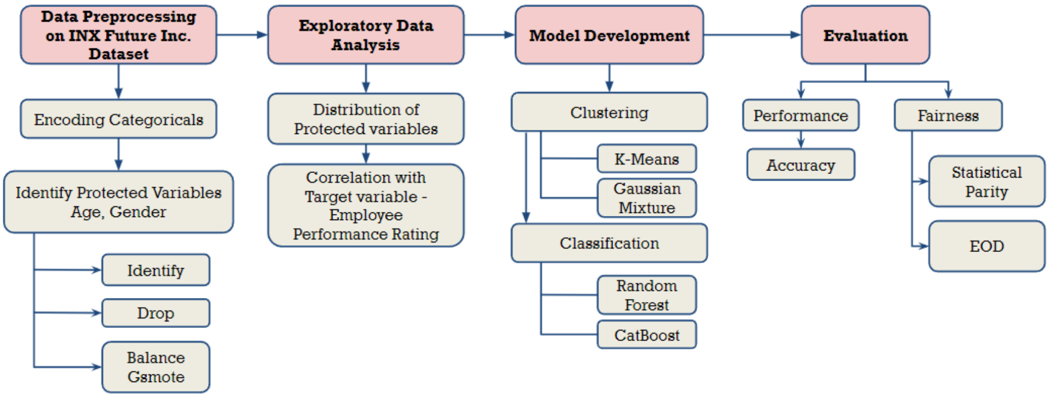

Figure 1: An overview of the methodology

cannot be ignored. While algorithms help improve human resource management efficiency by 84 %, they also introduce ethical and fairness issues and questions for organizations, which make it risky for employers to introduce AI to HRM (Park et al., 2022). A study recently discovered that algorithms assigned to human-centric decision-making tasks, such as recruitment and performance appraisals, are viewed by people as less trustworthy, less fair, and more negative (Lee, 2018). Moreover, bias is hard to measure as the perspective of fairness might be different from the employers' and employees' perspectives (Park et al., 2021). Park and co-authors also studied why people resisted AI in HRM, and they found that emotional, mental, bias, social, manipulation, and privacy factors can be reasons that hinder successful AI-based HRM. Some of these problems can be solved when designing an algorithm by introducing transparency and interpretability of the algorithm to narrow the employee's knowledge gap in implementing AI to perform employee evaluations.

In summary, our paper faces the complex task of examining algorithms and fairness concurrently. Although we can apply mathematical measures, like the confusion matrix and accuracy of prediction, to gauge fairness, this approach may not be wholly inclusive of other factors, such as the fairness perspective, and also contribute to what is fair in different circumstances.

4 Methodology

As shown in Figure 1, the methodology begins with the collection of employee performance data from Kaggle. Protected variables, such as age and gender, are identified and categorized to maintain the integrity of the analysis. The dataset undergoes preprocessing to address missing values and transform data into usable, standardized formats for consistency and comparability.

Next, in the exploratory data analysis (EDA) phase, we aim to gain a deep understanding of the dataset's characteristics. The analysis involves identifying patterns, detecting anomalies, and recognizing inconsistencies that could formulate hypotheses for further study.

For model development, We decided to use two different clustering algorithms, including K-means and Gaussian Mixture Model(GMM), to uncover hidden patterns and groupings in employee data. In K-means, the Silhouette score will be used to optimize the number of clustering. Moreover, the GMM is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. Unlike K-means, which assigns each data point to a single cluster, GMM assigns a probability (soft assignment) to each point belonging to each cluster. This method might help us identify unexpected relationships between employee characteristics and their performance ratings.

Furthermore, we use the clustering result as a feature of our prediction algorithm. We plan to construct supervised learning models, including Random Forest and CatBoost, for employee performance prediction due to their ability to model non-linear relationships and high predictive performance. By integrating specialized analyses and fairness-focused machine learning models, we want to make our evaluation process more equitable and transparent.

After the model is developed, we utilize accuracy to evaluate the accuracy of the AI models. We will compare accuracy metrics before and after applying bias mitigation strategies to assess how these strategies affect the model's accuracy. This comparison will help us understand the trade-offs between model accuracy and fairness.

For fairness assessment, we will use fairness metrics, including statistical parity and equal opportunity difference(EOD), to quantify bias in the models. These metrics help understand how the model's predictions may be biased towards or against particular groups based on age and gender. By comparing these fairness metrics before and after applying different bias mitigation strategies, we can evaluate the effectiveness of each technique in reducing bias. This contributes to identifying the most effective methods for promoting fairness in AI models.

In implementing this plan, our priority is to make the AI models accurate and fair, thereby promoting equity and trust in AI-based employee evaluations.

5 Results and Discussion

5.1 Exploratory Data Analysis

We explored the data through univariate and bivariate analyses to uncover insights and potential biases, particularly regarding gender and age. We focus on understanding the dataset's implications for predictive modeling and organizational dynamics. We utilized bar plots for categorical variables and examined correlations for numerical ones. We present the most insightful plots derived from our analysis.

5.1.1 Age Distribution

The dataset displays a right-skewed distribution of age, indicating a higher concentration of individuals aged 20-40, potentially biasing model outcomes (Figure 2). This skew persists across various performance ratings, notably in the 20-40 age group, suggesting a biased representation favoring younger participants (Figure 3).

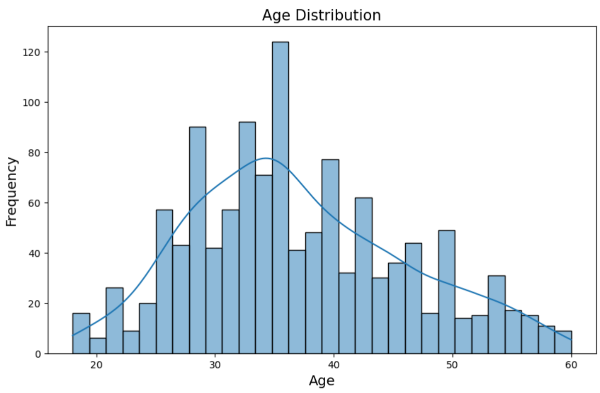

Figure 2: Distribution of Age in the INX Future Inc Dataset

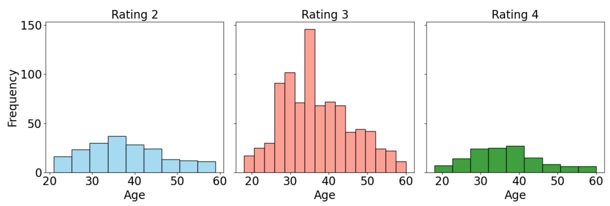

Figure 3: Distribution of Age across the various performance ratings in the INX Future Inc Dataset

5.1.2 Gender Disparity

Performance ratings reveal a gender bias, with males dominating the dataset across all performance ratings 2, 3, and 4 (Figure 4).

Addressing these biases is imperative for equitable modeling outcomes and fostering an inclusive organizational culture.

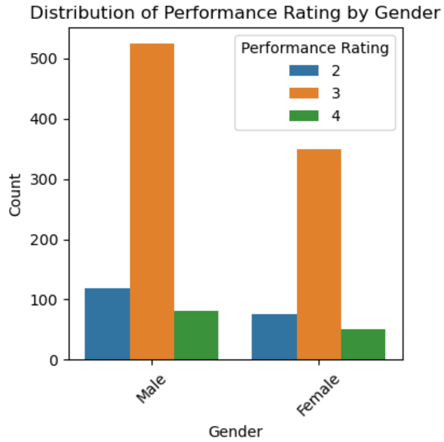

Figure 4: Bar plot of Gender distribution across various performance ratings in the INX Future Inc Dataset

5.1.3 Correlation

Referring to the correlation plot (Figure 5), the highest correlation with performance rating is observed for EmpEnvironmentSatisfaction (0.395561) and EmpLastSalaryHikePercent (0.333722), indicating a positive correlation. Conversely, YearsSinceLastPromotion (-0.167629), ExperienceYearsInCurrentRole (-0.147638), and YearsWithCurrManager (-0.122313) exhibit negative correlations, suggesting that longer durations since the last promotion or in the current role are associated with lower performance ratings.

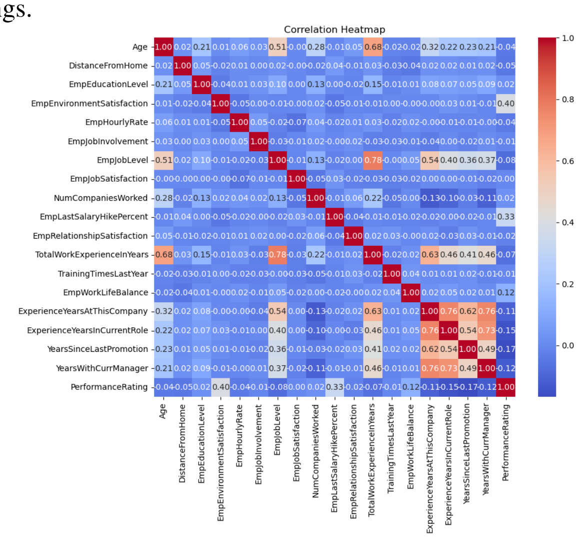

Figure 5: Correlation of the numerical features in the INX Inc Dataset

The positive correlation between performance rating and EmpEnvironmentSatisfaction implies that higher satisfaction at work correlates with better ratings, raising concerns about subjective influences on evaluations. Conversely, negative correlations with YearsSinceLastPromotion, ExperienceYearsInCurrentRole, and YearsWithCurrManager suggest biases against long-tenured employees, potentially due to career stagnation or biases favoring newer staff.

5.2 Resampling

As we observed in exploratory data analysis, the data is not uniformly distributed among the demographics such as age and gender. This can make the model biased for the category with a higher population. To resolve this issue and make the model fair, we resampled the dataset using the G-smote algorithm[1]. Figure 6 shows the distribution of age in the resampled dataset, which is balanced in comparison to the original dataset in Figure 2. Similarly, Figure 7 shows the distribution of gender in the resampled dataset, which is balanced in comparison to the original dataset in Figure 4

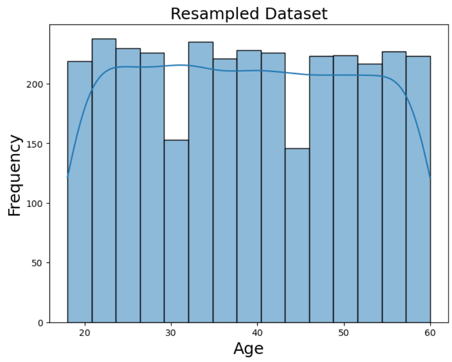

Figure 6: Bar plot of age distribution across various performance ratings in the resampled dataset

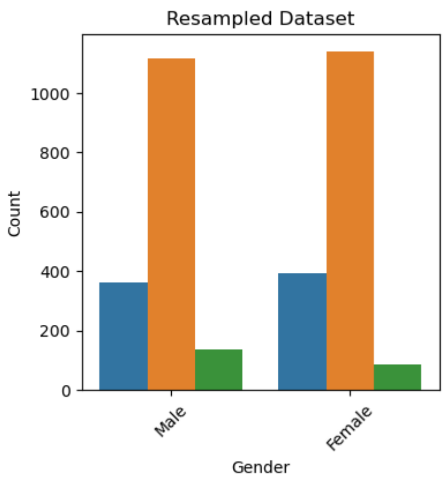

Figure 7: Bar plot of gender distribution across various performance ratings in the resampled dataset

5.3 Clustering

According to the K-means analysis result, the optimal value of K by Silhouette score is 2 for both original and G-smote datasets. The K-means clustering result is shown in Figure 8 and will be further used in the performance prediction algorithm.

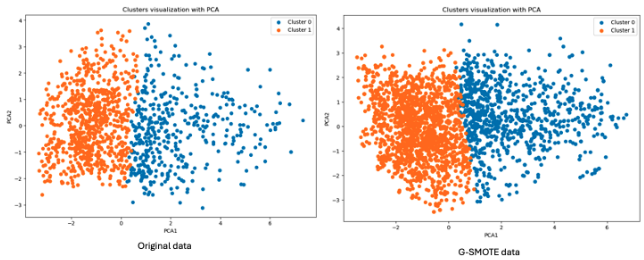

Figure 8: K-mean clustering for original and G-smote data respectively

The clustering result from GMM with original data and G-smote data is shown in Figure 9 and will be further used in the performance prediction algorithm.

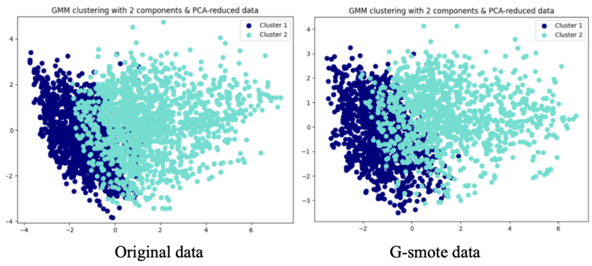

Figure 9: GMM clustering for original and G-smote data respectively

5.4 Algorithm

As we aim to maximize both accuracy and fairness for the employee performance algorithm, we constructed Random Forest and CatBoost models with the original dataset and G-smote dataset with different features as shown in Table 1, so that we can see the affect of G-smote and clustering clearly. Eventually, we got all sixteen models.

Table 1: Sixteen different models were built

Model	Dataset	Protected Variables	Clustering
Random Forest	Original	Not dropped	with K-mean
			without K-mean
		Dropped	with K-mean
			without K-mean
	G-smote	Not dropped	with K-mean
			without K-mean
		Dropped	with K-mean
			without K-mean
CatBoost	Original	Not dropped	with GMM
			without GMM
		Dropped	with GMM
			without GMM
	G-smote	Not dropped	with GMM
			without GMM
		Dropped	with GMM
			without GMM

The result of Random forest is shown below in Table 2. With original data, all cases provided promising results in terms of prediction accuracy. Comparing random forest without and with K-mean, the result shows that K-mean clustering does not significantly improve the accuracy of random forests as the accuracy only increases by 0.8 %.

Furthermore, the accuracy of the random forest algorithm with protected and without protected features is the same at 92.2 %, meaning that the inclusion or exclusion of protected features does not affect the model's accuracy. There are many reasons why accuracy remains the same, regardless of whether protected features are included. One of them is that the protected features may not be providing additional information useful for predictions beyond what is already captured by the other features. From a fairness perspective, achieving similar accuracy without using protected features is desirable. It suggests that the model can perform well without potentially introducing bias associated with those protected features.

With G-smote data, it is shown that the accuracy of all models is improving. However, K-mean decreases the accuracy of the model slightly. The best-performing model is a random forest without K-mean and dropping age and gender.

Table 2: Accuracy of Random Forest Algorithm

Random Forest	Accuracy original data	Accuracy G-smote data
without k-mean	0.914	0.939
with k-mean	0.922	0.938
without k-mean & drop age and gender	0.919	0.940
with k-mean & drop age and gender	0.922	0.939

The results for CatBoost model with original data and G-smote data are demonstrated in Table 3. With original data by introducing GMM cluster the model showed an improvement of 0.2 % and by removing protected features the accuracy dropped a little. Which shows that CatBoost model performance is taking the information from protected features. In-case of G-smote dataset, CatBoost model without GMM gave best accuracy of 94.7 %. By introducing the clustering, there is an accuracy drop of 0.4 %. But CatBoost with clustering showed an improved performance of 0.3 % by dropping the protected features. It is giving an accuracy of 94.6 %.

Similar to Random Forest model, CatBoost is also giving best accuracy with G-smote data. Even though the CatBoost that includes protected features might give higher accuracy, but it introduces the potential for bias within the model's predictions. The decision to remove protected features, despite a slight decrease in accuracy, reflects a prioritization of fairness and bias mitigation over maximal accuracy.

Table 3: Accuracy of Catboost Algorithm

CatBoost	Accuracy original data	Accuracy G-smote data
without GMM	0.919	0.947
with GMM	0.921	0.943
without GMM & drop age and gender	0.912	0.945
with GMM & drop age and gender	0.913	0.946

5.5 Fairness Aspect

For random forest, when considering statistical parity for the age group in Table 4, all models using the G-smote dataset perform worse than the original dataset for all performance ratings. On the one hand, with the original dataset, the K-mean doesn't affect statistical parity for the age group across all performance ratings, suggesting that the clusters found by K-means do not correspond to significant variations in age groups within different performance ratings. Also, dropping age and gender improves the metric for performance rating 2. On the other hand, with G-smote data, only the random forest model without K-mean and dropping age and gender improve statistical parity.

Table 4: Statistical Parity for Age for Random Forest

Statistical Parity for Age							
Random Forest	Age group	Performance rate 2		Performance rate 3		Performance rate 4	
		Original	G-smote	Original	G-smote	Original	G-smote
Without k-mean	0	-0.017	-0.074	0.016	0.079	0.000	-0.005
	1	0.028	0.079	-0.027	-0.084	-0.001	0.005
With k-mean	0	-0.017	-0.075	0.016	0.080	0.001	-0.005
	1	0.028	0.080	-0.027	-0.085	-0.001	0.006
Without k-mean & drop age and gender	0	-0.014	-0.072	0.017	0.077	-0.003	-0.004
	1	0.023	0.077	-0.029	-0.081	0.005	0.004
With k-mean & drop age and gender	0	-0.014	-0.078	0.010	0.083	0.004	-0.005
	1	0.023	0.083	-0.017	-0.088	-0.007	0.005

When considering statistical parity for the gender group, all models using the G-smote dataset perform better than the original dataset for all performance ratings, but rating 4. On the one hand, with the original dataset, the K-mean doesn't affect statistical parity for the gender group across all performance ratings. Also, dropping age and gender improves the metric for performance rating two but gives varying results for another performance rating. On the other hand, with G-smote data, dropping age and gender improves the metric for all performance ratings, except for performance rating 4 for random forest with K-mean.

Table 5: Statistical Parity for Gender for Random Forest

Statistical Parity for Gender							
Random Forest	Gender group	Performance rate 2		Performance rate 3		Performance rate 4	
		Original	G-smote	Original	G-smote	Original	G-smote
Without k-mean	0	-0.030	0.014	0.042	-0.003	-0.012	-0.012
	1	0.019	-0.014	-0.027	0.003	0.008	0.011
With k-mean	0	-0.030	0.015	0.042	-0.004	-0.012	-0.012
	1	0.019	-0.015	-0.027	0.004	0.008	0.011
Without k-mean & drop age and gender	0	-0.027	0.008	0.041	0.002	-0.014	-0.010
	1	0.0173	-0.008	-0.026	-0.002	0.009	0.010
With k-mean & drop age and gender	0	-0.027	0.014	0.031	-0.003	-0.004	-0.012
	1	0.017	-0.013	-0.020	0.003	0.002	0.011

Considering EOD for age group in Table 6, for Performance rates 2 and 3, the application of G-smote generally results in lower EOD values, implying improved fairness across different configurations. However, for Performance rate 4, the results are mixed, with some configurations showing improved fairness and others indicating increased bias when G-smote is applied. It is notable that dropping age and gender from the model seems to yield lower EOD values, especially when G-smote is not applied, which may imply that these features contribute to age-related bias in the model predictions. However, when combined with G-smote, dropping these features can sometimes lead to higher EOD values, as seen in Performance rate 4, suggesting that the interaction between feature selection and synthetic data generation needs careful consideration.

Table 6: Equal Opportunity difference for Age for Random Forest

Equal Opportunity for Age						
Random Forest	Performance rate 2		Performance rate 3		Performance rate 4	
	Original	G-smote	Original	G-smote	Original	G-smote
Without k-mean	0.116	0.062	0.027	0.003	0.165	0.156
With k-mean	0.116	0.052	0.027	0.004	0.165	0.081
Without k-mean & drop age and gender	0.071	0.047	0.027	0.005	0.131	0.123
With k-mean & drop age and gender	0.071	0.080	0.017	0.002	0.025	0.156

Applying G-smote to Random Forest models generally leads to improved fairness in terms of EOD for gender, particularly for performance ratings 2 and 3. This suggests G-smote's effectiveness in reducing gender bias at these levels. However, for the higher Performance rate 4, the impact of G-smote is less straightforward, with some models showing an increase in gender bias. Notably, models that include K-means clustering maintain the same level of EOD for gender when G-smote is applied, while models that exclude age and gender features demonstrate a significant increase in Opportunity Differences for gender, indicating a negative impact on fairness. This detailed analysis underscores the complex interplay between algorithmic interventions like G-smote and feature selection strategies.

Table 7: Equal Opportunity Difference for Gender for Random Forest

Equal Opportunity for Gender						
Random Forest	Performance rate 2		Performance rate 3		Performance rate 4	
	Original	G-smote	Original	G-smote	Original	G-smote
Without k-mean	0.093	0.061	0.041	0.004	0.059	0.081
With k-mean	0.093	0.052	0.041	0.004	0.059	0.081
Without k-mean & drop age and gender	0.125	0.086	0.041	0.005	0.131	0.137
With k-mean & drop age and gender	0.125	0.060	0.034	0.004	0.236	0.081

For CatBoost, when considering statistical parity for age in Table 8, all models using the G-smote dataset perform worse than the original dataset for performance ratings 2 and 3, whereas for performance rating 4 G-smote outperformed the original dataset. Among all models CatBoost with clustering and removing protected features is giving the best results for statistical parity for age.

Table 8: Statistical Parity for Age for catboost

Statistical Parity for Age							
Catboost	Age group	Performance rate 2		Performance rate 3		Performance rate 4	
		Original	G-smote	Original	G-smote	Original	G-smote
Without GMM	0	-0.063	-0.081	0.044	0.084	0.019	-0.003
	1	0.068	0.086	-0.047	-0.089	-0.020	0.003
With GMM	0	-0.064	-0.081	0.044	0.081	0.02	0.001
	1	0.069	0.086	-0.047	-0.086	-0.021	-0.001
Without GMM & drop age and gender	0	-0.066	-0.072	0.047	0.076	0.019	-0.003
	1	0.071	0.077	-0.050	-0.080	-0.020	0.004
With GMM & drop age and gender	0	-0.064	-0.074	0.044	0.075	0.020	-0.001
	1	0.069	0.079	-0.047	-0.079	-0.021	0.001

When considering statistical parity for gender in Table 9, all models using the G-smote dataset outperformed the original dataset for performance ratings 2 and 3, whereas for performance rating 4 original dataset outperformed the G-smote dataset. Among all models CatBoost without clustering and removing protected features is giving the best results for statistical parity for gender.

Table 9: Statistical Parity for Gender for Catboost

Statistical Parity for Gender							
Catboost	Gender group	Performance rate 2		Performance rate 3		Performance rate 4	
		Original	G-smote	Original	G-smote	Original	G-smote
Without GMM	0	0.027	0.020	-0.013	-0.005	-0.014	-0.014
	1	-0.027	-0.019	0.012	0.005	0.014	0.014
With GMM	0	0.024	0.020	-0.015	-0.009	-0.009	-0.011
	1	-0.024	-0.019	0.015	0.008	0.009	0.011
Without GMM & drop age and gender	0	0.022	0.006	-0.009	0.007	-0.012	-0.013
	1	-0.022	-0.006	0.009	-0.007	0.012	0.013
With GMM & drop age and	0	0.024	0.010	-0.013	0.004	-0.011	-0.014
gender							
	1	-0.024	-0.009	0.012	-0.004	0.011	0.013

When considering EOD for age in Table 10, all models using the G-smote dataset outperformed the original dataset for performance ratings 3 and 4. However, for performance rating 4, the original dataset outperformed the G-smote dataset. Among all models, CatBoost with clustering and removing protected features is giving the best results for EOD for age.

Table 10: Equal Opportunity Difference for Age for Catboost

Equal Opportunity Difference for Age						
Catboost	Performance rate 2		Performance rate 3		Performance rate 4	
	Original	G-smote	Original	G-smote	Original	G-smote
Without GMM	0.046	0.085	0.023	0.013	0.349	0.026
With GMM	0.026	0.107	0.026	0.005	0.321	0.042
Without GMM & drop age and gender	0.039	0.034	0.029	0.005	0.222	0.059
With GMM & drop age and gender	0.043	0.068	0.021	0.004	0.25	0.007

In Table 11, when assessing equal opportunity difference for gender, all models, except the CatBoost model without GMM that use G-smote dataset, show superior performance on the original dataset for a performance rating of 4. For a performance rating 3, both the CatBoost models without and with GMM, and without protected features achieved identical results on both the original and G-smote datasets. However, the CatBoost model with GMM trained on the original dataset outperformed the one using the G-smote dataset for this performance rating. Specifically, the CatBoost model without GMM and without protected features on the G-smote dataset significantly surpassed its counterpart on the original dataset. For a performance rating of 2, the original dataset performed better than the G-smote dataset. Overall, among all the models evaluated, the CatBoost model without GMM was the most effective.

Table 11: Equal Opportunity Difference for Gender for Catboost

Equal Opportunity Difference for Gender						
Catboost	Performance rate 2		Performance rate 3		Performance rate 4	
	Original	G-smote	Original	G-smote	Original	G-smote
Without GMM	0.001	0.037	0.007	0.007	0.103	0.157
With GMM	0.019	0.036	0.010	0.011	0.316	0.206
Without GMM & drop age and gender	0.028	0.101	0.013	0.008	0.224	0.181
With GMM & drop age and gender	0.023	0.069	0.004	0.004	0.195	0.133

link to our source code: https://github.com/ThanawanL/DSCI531_employee_evaluation_final_project.git

6 Conclusion and Future Work

In our analysis journey, we conducted Exploratory Data Analysis to uncover key features, resampled the dataset, and built two algorithms for modeling and used statistical parity and equal opportunity difference as fairness measures.

Regarding random forest algorithm, for age-based statistical parity, G-smote generally worsens disparities, except when features like age and gender are dropped, where it improves fairness for lower performance ratings but not uniformly across all metrics. For gender-based statistical parity, G-smote improves outcomes across almost all performance ratings, indicating its potential to mitigate gender bias, albeit with less effectiveness at the highest performance rating. The implementation of K-means clustering shows a neutral effect on fairness for both age and gender groups, suggesting that it does not capture significant age or gender-related variations.

The CatBoost model showed varied performance across different fairness metrics. For age-based analysis, the models using the G-smote dataset generally performed worse in statistical parity for lower performance ratings but showed improvements for higher ratings. Interestingly, when protected features were removed, there was an observable enhancement in fairness. For gender-based fairness, the results were more promising with the G-smote dataset improving outcomes significantly across almost all performance ratings. However, the highest performance rating still presented challenges. Among the variations tested, the CatBoost model with GMM clustering and without protected features consistently delivered strong performances, indicating its effectiveness in managing biases while maintaining robust model functionality. These findings highlight the capability of CatBoost in navigating the complexities of algorithmic fairness, especially when tailored through strategic data preprocessing and removing protected features.

Future work in this area could focus on several key aspects. Firstly, exploring more sophisticated bias mitigation techniques that take into account the complex interactions between various features and model predictions could be beneficial. Since dropping protected variables like age and gender has shown mixed results, future studies might investigate the nuanced effects of such variables within different performance ratings more deeply. Secondly, expanding the analysis to include other forms of bias beyond age and gender could provide a more comprehensive view of fairness. This could include socioeconomic status, departmental bias, or other forms of indirect discrimination that may affect model predictions. Lastly, exploring alternative machine learning models or ensemble methods that may offer different advantages in balancing accuracy and fairness is another worthwhile direction for research. Comparing such models could yield insights into the types of algorithms that are better suited for fair performance evaluations.

References

[1] G. Douzas and F. Bação. Geometric smote a geometrically enhanced drop-in replacement for smote. *Information Sciences*, 501, 06 2019.

[2] A. S. Lather, R. Malhotra, P. Saloni, P. Singh, and S. Mittal. Prediction of employee performance using machine learning techniques. In *AISS '19: Proceedings of the 1st International Conference on Advanced Information Science and System*, New York, NY, USA, 2019. Association for Computing Machinery.

[3] M. K. Lee. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. SageJournals, 2018.

[4] R. Noe. Human resources management: Gaining a competitive advantage. New York, MA, USA, 2006. McGraw-Hill Education.

[5] H. Park, D. Ahn, K. Hosanagar, and J. Lee. Human-ai interaction in human resource management: Understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[6] H. Park, D. Ahn, K. Hosanagar, and J. Lee. Designing fair ai in human resource management: Understanding tensions surrounding algorithmic evaluation and envisioning stakeholder-centered solutions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[7] P. Sujatha and R. Dhivya. Ensemble learning framework to predict the employee performance. Raipur, India, 2022. IEEE.