Towards a Novel Visual Evaluation of Algorithmic Bias: Insights on the Italian Academic System

Francesca Ciccarelli
francesca.ciccarelli@student.univaq.it
Università degli Studi dell'Aquila
L'Aquila, Italy

Andrea D'Angelo
andrea.dangelo6@graduate.univaq.it
Università degli Studi dell'Aquila
L'Aquila, Italy

Giovanni Stilo
giovanni.stilo@univaq.it
Università degli Studi dell'Aquila
L'Aquila, Italy

ABSTRACT

Bias in real-world applications and machine-learning systems leads to inequitable outcomes and perpetuates disparities. Traditional methods often fail to capture bias complexities fully. We introduce a novel use of the ROC curve to analyze classifier performance across subgroups, offering a more detailed understanding of bias. Validated through a case study on the Italian academic system, our approach effectively evaluates gender disparities in career progression. Our contributions include an innovative ROC curve application, practical validation, and a framework for enhancing fairness and inclusivity in decision-making processes.

CCS CONCEPTS

• **Computing methodologies → Machine learning approaches**; • **Applied computing →** *Law, social and behavioral sciences.*

KEYWORDS

Data Analytics, Gender bias, Machine Learning

ACM Reference Format:
Francesca Ciccarelli, Andrea D'Angelo, and Giovanni Stilo. 2024. Towards a Novel Visual Evaluation of Algorithmic Bias: Insights on the Italian Academic System. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'KDD).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Bias is a significant issue with implications both in real-world applications and machine learning systems. Whether in human judgment or machine learning algorithms, bias can lead to inequitable outcomes and perpetuate existing disparities. In real-world contexts, biases can manifest in various forms, such as racial, gender, or socioeconomic biases, influencing hiring, lending, and law enforcement [3]. In machine learning, biases in training data or algorithmic design can result in models unfairly disadvantaging certain groups [26]. Addressing these biases is critical for developing fair and ethical AI systems that align with societal values [19] as those stated by the UN Sustainable Development Goals 5 and 10 [33].

The existing quantitative methods, such as accuracy metrics or confusion matrices, are insufficient to fully capture the complexity of the problem. They often fall short of revealing the nuanced ways in which bias can affect model performance [8]. These methods typically provide aggregate measures that overlook the differential impact on subgroups. For instance, a model might perform well overall but still exhibit significant disparities in performance across different demographic groups. Hence, there is a need for more tools and methodologies that can capture the multifaceted nature of bias and its implications in machine learning [9].

Therefore, in this work, we employ Receiver operating characteristic (ROC) curves in an innovative manner to enhance the understanding of the algorithmic foundations of a classifier. The ROC curve is a well-established tool for evaluating the performance of binary classifiers, plotting the true positive rate against the false positive rate at various threshold settings [15]. Our intuition is that by examining the ROC curves across different subgroups, we can identify and analyze disparities in model performance. This approach provides a more granular understanding of how bias manifests and offers a pathway to developing more equitable machine learning models [10].

As a case study, we analyzed data from the Italian academic system, focusing on the progression from researcher to associate professor and from associate to full professor. This analysis aims to evaluate the effectiveness of our innovative use of the ROC curve in detecting and addressing biases in real-world scenarios. The Italian academic system provides a pertinent case study due to its well-documented gender disparities and other biases in career progression (see Section 2). By examining the transition rates between academic ranks, we can identify patterns of inequity that may be influenced by systemic biases. Moreover, we show how our approach can also be adopted to evaluate the impact of a debiaser (namely, DEMV [14]) in a more granular way. This practical case highlights the utility of our approach as an empirical tool for evaluating and addressing biases. Our contributions can be summarised as follows:

- Framework for Equity: we depict a methodological framework to critically assess and improve their decision-making processes, promoting fairness and inclusivity;
- Novel use of ROC Curve: we introduced an innovative application of the ROC curve to detect and analyze algorithmic biases, providing a more granular understanding of bias manifestation in machine learning classifiers;
- Real-World Application: we validated the practical utility of the proposed approach by applying it to a real-world scenario, showing its potential as a tool for evaluating and addressing biases in various contexts.

Conference acronym 'KDD, June 25–05, 2024, Barcellona, Spain

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/10.1145/nnnnnnn.nnnnnnn

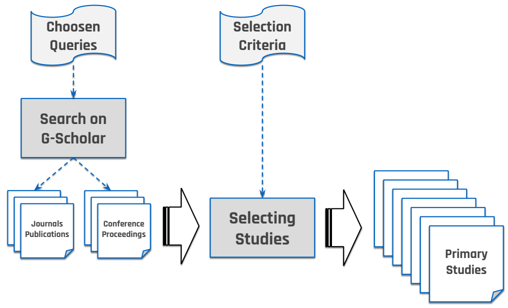


Figure 1: Literature Review Workflow

The paper is structured as follows: in Section 2, we go over the technical-related works and the studies on bias in academia. In Section 3, we detail our proposed methodology for the visual evaluation of algorithmic bias. Then, in Section 4, we specify our experimental settings for the experiments we conducted and for which we visualize results in Section 5. Lastly, in Section 6, we draw our data-driven conclusions.

2 BACKGROUND KNOWLEDGE AND RELATED WORK

In this section, we review related works, organized into two subfields. First, we discuss technical-related works and concepts. Then, we examine studies on bias in academic systems, which form the foundation of our case study.

Technical Related Works. There are several bias metrics commonly used to evaluate gender disparities, such as the *Demographic Parity*[13], *Equal Opportunity*[22], and *Disparate Impact*[16]. In this study, we adopt the Disparate Impact as a quantitative measure to measure the bias present in a dataset since it does not need to rely on any predictor. This makes it a robust and versatile metric for identifying and addressing bias in various contexts. Several methods for decreasing the intrinsic bias of a dataset exist. In this work, we make use of DEMV [14], a debiaser for multiple variables that operates on the dataset rather than learning models.

The *ROC curve*[15] provides a visual tool to assess the performance of machine learning models - particularly in binary classification tasks - enabling the data scientist to make informed decisions on model selection and threshold settings.

Related Works on Bias in Academics. To better understand the current academic landscape, we conducted a Systematic literature review focused on those works that lie in the intersection of algorithmic bias and Learning Systems. To do so, we follow the workflow depicted in Fig. 1 by collecting works obtained by searching Google Scholar for the queries reported in Figure 2. We focus primarily on Italian works because they operate within our study's career advancement framework. However, for completeness, we also include studies on foreign educational systems to broaden our perspective. Specifically, we include articles related to the recruitment, promotion, and productivity level of academic staff, i.e., full professors, associate professors, and researchers. We do not include articles pertaining to specific faculties or the gender bias present in the general working world.

- **Q1**: "Gender bias in academic recruitment"
- **Q2**: "Italian academia gender discrimination"
- **Q3**: "Women's success in faculty recruitment"
- **Q4**: "Women's faculty equity in academia"
- **Q5**: "Gender in career advancements in Italian universities"
- **Q6**: "Academic promotions gender discrimination"
- **Q7**: "Gender bias in selection processes for professors"
- **Q8**: "Impact of bias in faculty recruitment"
- **Q9**: "Gender differences in the Italian academic system"
- **Q10**: "Gender discrimination and promotion in academia"
- **Q11**: "Impact of gender and family factors in productivity"
- **Q12**: "Female representation in academic institutions"

Figure 2: Queries adopted for the literature review.

After the collecting and selection phases, we classified the literature by focusing on four main aspects: *Reference Context*, *Purpose of the study*, *Used Data*, and the *Adopted Methodology*.
In the *Reference Context*, there are the works [1, 2, 5, 11, 17, 20, 25, 27, 28] which deal with problems inherent to gender bias in Italian university institutions. While, works as [4, 6, 7, 21, 23, 24, 32, 34] focus on the study of gender bias in foreign universities such as those in the Nordic regions, or Germany, Austria, Switzerland, United States, Australia and Scotland. And works as [18, 29–31] concern generic studies in the field that are not related to a specific university. The *Purpose of the studies* can be organised in two main categories. Those which analysing *recruitment* issue [2, 4, 6, 7, 20, 21, 28–31], and those interested in *career promotion*[6, 27, 32]. For the *Source Data*, there are papers that works on *public data* [7, 17, 18, 20, 21, 23, 28, 28, 29, 31] - mainly from the MIUR database - and those [1, 2, 4, 5, 25, 30, 32] that used *private data*. The core part of our review focused on understanding which are the typical *Adopted Methodologies*. There are works that use *Descriptive statistics* [5, 20, 21] which analyse the percentages of males and females across career stages and institutions, means, standard deviations, min, max, or comparisons using t-tests between men and women. We classified under the umbrella of *Statistical analysis* those works [28, 34] that investigate gender inequality using different types of regressions, such as OLS regressions, multiple logistic regressions, and multilevel logistic regressions. Lastly, there are those works [7, 31] in which *Qualitative* analyses were carried on through surveys, questionnaires, and interviews involving all levels of academic staff. Table 1 summarises the aforementioned classification in one place.
In summary, research consistently shows that gender biases disadvantage women in academia, negatively impacting their promotion, productivity, and recruitment negatively. Reforms have addressed this issue, but barriers remain despite women's comparable productivity.

3 FRAMEWORK FOR EQUITY

In this section, we detail the methodology that we propose as a Framework for Equity. The Framework for Equity consists of the following steps as depicted in Figure 3:

Selection of the Best Predictor. In the first phase, we rigorously evaluate (e.g., 10-fold cross-validation) various predictive models to

	Context	Purpose		Source Data	Methods		
Paper	Italian	Recruitment	Promotion	Public	Statistical	Descriptive	Qualitative
[32]			X			X	X
[34]			X		X		X
[31]		X	X		X	X	X
[18]					X	X	X
[4]		X				X	
[1]	X			X		X	
[21]		X	X		X	X	
[27]	X		X	X	X	X	
[2, 28]	X	X		X	X	X	
[11, 17]	X		X	X		X	
[25]	X		X	X	X		
[6, 29]		X	X				X
[30]		X				X	X
[20]	X	X	X	X		X	
[5]	X		X		X	X	
[7]		X			X		X
[23]			X		X	X	
[24]			X			X	

Table 1: Summary of the Analysis of the Literature.

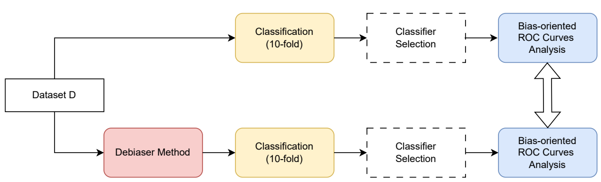


Figure 3: Overview of the Framework for Equity

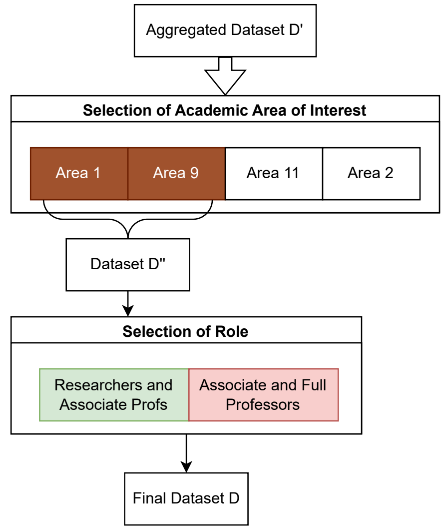


Figure 4: Processing pipeline of the dataset.

identify the best-performing algorithm for the specific use case. This involves assessing models based on accuracy, precision, recall, and other relevant metrics, such as the ROC curve. The chosen predictor should demonstrate high overall performance while applying to the context of the conducted study.

Visual Evaluation of Biases. After selecting the best predictor, we visually evaluate biases present in the predictions. This is achieved by creating specialized versions of the ROC curve, segmented by sensitive variables such as gender, race, or socioeconomic status. These subgroup-specific ROC curves are plotted alongside the original ROC curve, enabling the researcher to identify which and how minority groups underperform, indicating biased behavior from the model. This approach works best for binary-sensitive variables but can be easily extended to multi-class cases using the one-vs-all approach (training a binary classifier for each class).

Debiasing. In the third phase, we apply debiasing techniques to the dataset. This can involve reweighting, resampling, or adjusting the decision thresholds to reduce bias. The goal is to mitigate the identified disparities without significantly compromising the model's overall performance.

Validation of the Best Predictor. According to the adopted debiasing techniques, it may be necessary to validate the predictor model to confirm that it performs best. If the prior results are not confirmed, it will be necessary to reapply the methodology used in the selection phase of the best predictor. This might involve re-evaluating different algorithms, tuning hyperparameters, or exploring alternative models to ensure the predictor achieves high accuracy and consistent performance.

Visual Analysis and Comparison. Finally, we conduct a comprehensive visual analysis and compare the results obtained after debiasing. This involves plotting the ROC curves for each subgroup before and after applying debiasing techniques. By comparing these curves, we validate the improvements made in fairness and ensure that the model's performance has not degraded consistently. This step is crucial for demonstrating that our debiasing efforts have led to a more equitable model while maintaining its predictive accuracy and reliability.

By following the proposed Framework for Equity, it is possible to systematically visually identify, address, and validate biases in predictive models, promoting fairness and inclusivity in decision-making processes.

4 EXPERIMENTAL SETTINGS

This section details the experimental settings encompassing the dataset, predictor models, and debiaser involved in our study.

Dataset. The dataset used in this study is an aggregated collection of academic records sourced from Scopus, the MIUR website, and manually integrated with Google Scholar. This dataset provides a comprehensive view of academic productivity and career progression from 2015 to 2022. It includes temporal data on publication citations, academic roles, and gender for each scholar.

Figure 4 shows the preprocessing pipeline we applied to the dataset for our use case.

In this study, the focus is only on Areas 1 and 9 of the Italian minister of education (*"Ministero dell'Istruzione e del Merito - MIUR"*) [12] scientific areas classification, which broadly refers to Science, technology, engineering, and mathematics. From this further filtering, a dataset D'' was obtained. Finally, to binarize the labels, we split them into two different datasets on the promotions from Researcher to Associate Professor and from Associate to Full Professor, respectively. For simplicity, from now on, we refer to *DR*

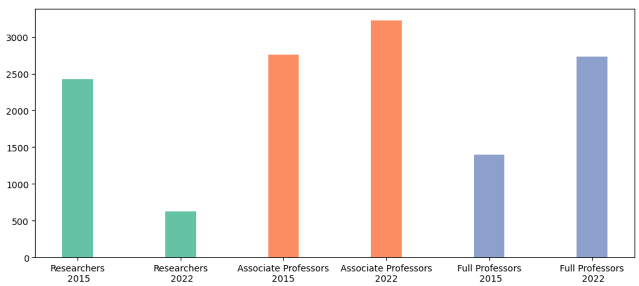


Figure 5: Number of researchers, associate professors, and full professors in the years 2015 and 2022.

as the dataset that contains the progression from Researcher to Associate professor. We refer to *DA* as the dataset that contains the progression from Associate to Full Professor.

To avoid contamination from other academic systems, we exclusively consider scholars who joined the Italian academia no later than 2015 and remained consistently within it through 2022. For this reason, the number of Researchers is bound to decrease, while the number of Associate and Full Professors will increase on a yearly basis. In particular, Figure 5 shows the cardinality of Researchers, Associate Professors, and Full Professors in our dataset for 2015 and 2022.

Specifically, there were *628* researchers, *3222* Associate Professors and *2737* Full Professors in 2022. In the year 2015, the presence of researchers was higher than in 2022 since many progressed to associate professors and fewer to full professors. This year, in fact, the number of researchers was *2428*, the number of associate professors was *2758*, and the number of full professors was *1401*. However, the number of associate professors is still higher than that of full professors due also to the greater difficulty of career progression.

Figure 6 illustrates the career progression of researchers and professors from 2015 to 2022. The orange section represents the *1664* researchers who advanced to associate professors by 2022, and the light green section represents the *137* researchers who progressed to full professors in the same year. Additionally, the dark orange section shows the associate professors who remained in their positions from 2015 to 2022. The green section highlights the *1201* associate professors from 2015 who advanced to full professors by 2022, joining the *1399* full professors already present, as indicated by the dark green section.

Researchers who have progressed in their careers from 2015 to 2022 amount to *1664*, while those who have not progressed are *627*.

The bar chart in Figure 7 shows the gender distribution of males and females across the three roles of researcher, associate professor, and full professor.

As expected, the number of males is much higher across all roles, making up roughly 70% of scholars in our dataset.

Employed predictor models. In the first step of our framework -the *Selection of the Best Predictor* - we consider five different classifier models as follows:

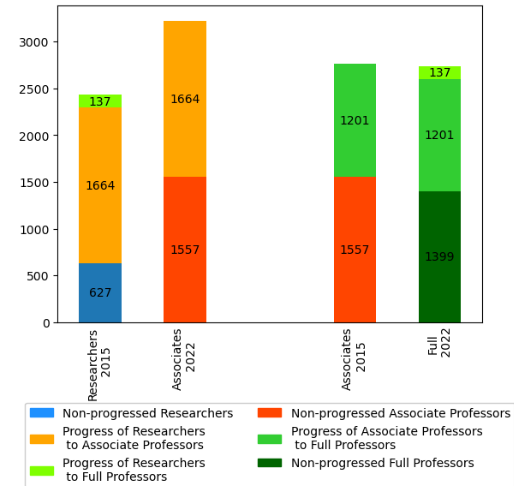


Figure 6: Number of researchers progressed to associate professors, and number of associate professors advanced to full professors from 2015 to 2022.

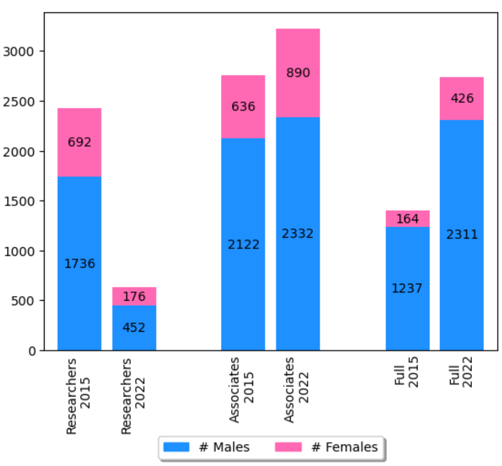


Figure 7: Bar chart on male and female gender distribution in the three roles of researcher, associate professor and full professor.

Multilayer Perceptron (MLP): a neural network with multiple layers used for classification. We used two network architectures, both of which had ReLU activation functions. The first one (MLP) uses 120, 100, 50, and 25 neurons in the four hidden layers. The second one (MLP2) uses 50, 25, and 12 neurons in the three hidden layers. Both the architectures were trained for 1000 epochs with the Adam optimizer.

Support Vector Machine (SVM): a classification algorithm that finds the best hyperplane to separate data into different classes.

eXtreme Gradient Boosting (XGBoost): a tree-based classification algorithm that aims to predict classes accurately.

Logistic Classifier: a simple yet effective algorithm for estimating the probability that a given input belongs to one of two classes.

Linear Classifier: an algorithm that uses a straight line, or a hyperplane in multidimensional spaces, to separate data into different classes.

All the results were computed using a K-fold cross-validation (k equal to 10), which is a standard method for estimating the performance of a machine learning algorithm on a dataset.
Once we have applied the different models to our analysis, we select the model with the highest AUC, so that we can then use it to visualize the ROC curves. Through the representation of the ROC curves, we aim to visualize the gender biases present, relating to the promotion from researcher to associate professor and from associate to full professor, as described in the *visualization* phase of our process.

Debiasing. In the *debiasing* phase of our process, we employ DEMV [14], a Debiaser for Multiple Variables, a method that mitigates unbalanced groups bias (i.e., bias caused by an unequal distribution of instances in the population), thus operating at the dataset level rather than on the models themselves. Since DEMV is model-agnostic, it works seamlessly with any of the classification models we used without needing to interfere with them individually. This allows for consistent application of bias mitigation across different models, ensuring that the debiasing process does not require model-specific adjustments or configurations.
The debiased dataset obtained after applying DEMV was then used to validate the best model, always chosen from those defined in section 4 with the highest AUC. The results obtained from the debiased dataset were then displayed and compared with the results relating to the original dataset.

5 RESULTS ANALYSIS

In this section we illustrate the results we obtained across all our experiments. Specifically, we first show the ROC curves obtained by applying the classification models on the dataset directly (refer to Figure 3). Then, we show the ROC curves obtained on the debiased datasets, and comment upon the differences.
The ROC curve line shows the trade-off between the true positive rate (sensitivity) and the false positive rate (specificity) across different threshold settings. A curve that bows toward the top-left corner indicates good performance while a straight diagonal line indicates a model with no discriminative power (i.e., equivalent to random guessing).

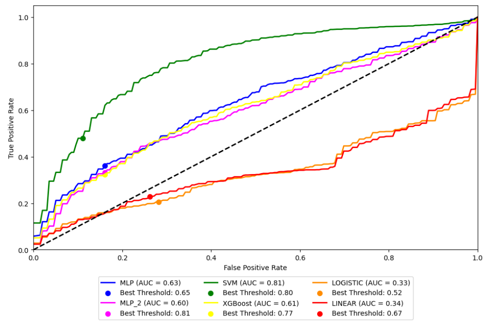
Analysis of ROC Curves of Original Biased Datasets. Figure 8a shows that SVM is the model with the highest AUC (AUC=0.81) and it was used to compare predictions with data for men and women.
Specifically, the red curve is for the model trained on the researchers' original dataset, without gender distinction. The pink and blue curves, on the other hand, represent the specialized model for males and females.
It is possible to see from figure 8c that the model performs well in all case studies, with better performance on the male gender, which indicates the unfairness embedded in the classifier. Notably, this result was not captured by the measured Disparate Impact (DI) of 1.00738, which indicates - on the opposite - a fair classifier.
The same analysis was performed for the dataset of Associate Professors (figure 8b) and, also in this case, the best model was found to be SVM (AUC=0.72).
Figure 8d shows a slightly better performance for female associate professors, but it is possible to notice that the AUC values are almost similar. This result is in line - even if it is not perfectly aligned - with the measured DI of 0.8116.

Analysis of the ROC Curves for the Debiased Datasets. Using the researchers' debiased dataset, it is possible to notice in figure 9a that the model with the best performance was XGBoost (AUC=0.84).
The comparison for the gender bias was now performed using the model XGBoost and the results obtained are shown in figure 9c.
For male researchers it is possible to note an excellent correspondence with respect to the ROC curve obtained without gender distinction, while the performance on the female gender is slightly lower but still has a good AUC value. Notably, also this result was not captured by the measured Disparate Impact (DI) of 1.0014, which indicates an almost perfect fair dataset.
The same analysis, performed on the debiased dataset related to associate professor, shows that the best model now is again the SVM (AUC=0.72); in figure 9b can be also seen that Logistic Classifier has the same AUC and XGBoost has more or less the same result.
For the comparison we decide to use the SVM model and the results are shown in figure 9d.
The model relating to associate professors shows slightly better performance for the female gender, with an AUC equal to 0.75. This result is in line with the measured DI of 0.9047.

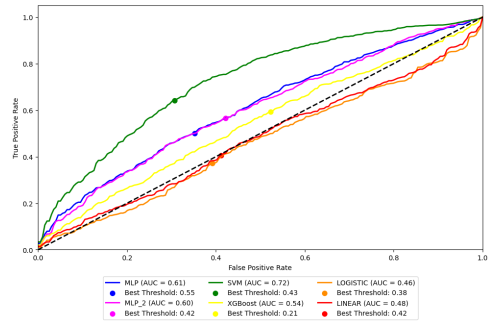
Comparison and Analysis of ROC Curves. The results from the two analyses indicate that the analysis conducted on the debiased dataset achieves equal or higher Area Under the Curve (AUC) values. This suggests a notable improvement in fairness, particularly evident in the researchers' data.

6 CONCLUSION AND FUTURE WORKS

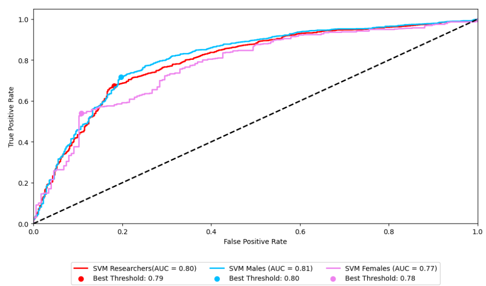
In this paper, we presented a novel method for using ROC (Receiver Operating Characteristic) curves to enhance the understanding of biases in classifier algorithms. We applied this approach to the Italian academic system, known for its gender disparities, to analyze transitions between academic ranks and evaluate the effectiveness of debiasing techniques.
Our analysis highlights the innovative use of ROC curves to detect biases in machine learning models. By examining these curves across different subgroups, we identified significant disparities in model performance that were not always evident through traditional metrics like Disparate Impact. The case study on the Italian academic system illustrated how gender biases could be uncovered and mitigated, showing improvements in fairness after applying debiasing techniques. This approach underscores the value of combining multiple evaluation methods to achieve a more nuanced and equitable assessment of classifier performance.
Future works include assessing the methodology on diverse datasets to test its robustness and comparing it to other bias metrics to highlight its strengths and pitfalls.



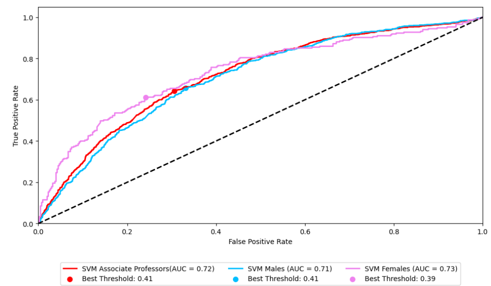
(a) ROC curves of the different models for the biased *DR*.



(b) ROC curves of the different models for the biased *DA*.

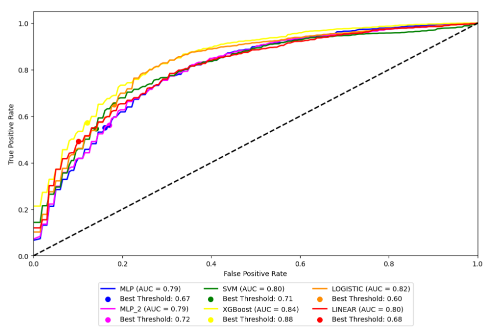


(c) Gender-specific ROC curves of the SVM on the biased *DR*.

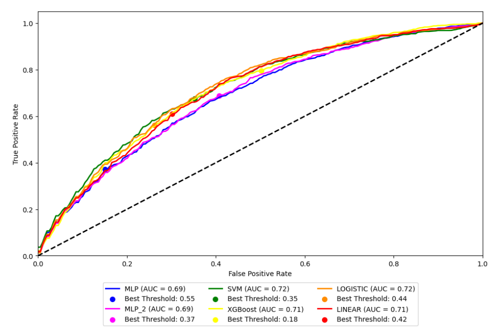


(d) Gender-specific ROC curves of the SVM on the biased *DA*.

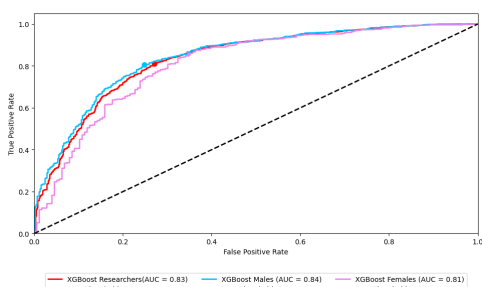
Figure 8: Comparison (best predictor and gender-oriented) of the original datasets without applying any debiasing technique.



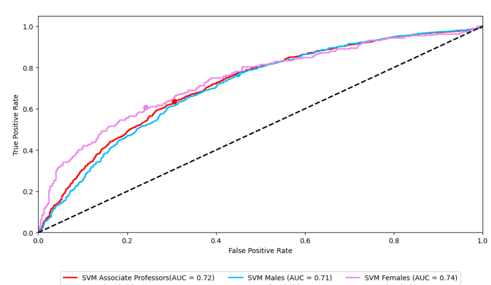
(a) ROC curves of the different models for the unbiased *DR*.



(b) ROC curves of the different models for the unbiased *DA*.



(c) Gender-specific ROC curves of the XGBoost on the unbiased *DR*.



(d) Gender-specific ROC curves of the SVM on the unbiased *DA*.

Figure 9: Comparison (best predictor and gender-oriented) of the datasets obtained by applying debiasing technique.

REFERENCES

[1] Giovanni Abramo, Ciriaco D'Angelo, and Alessandro Caprasecca. 2009. Gender Differences in Research Productivity: A Bibliometric Analysis of the Italian Academic System. *Scientometrics* 79 (06 2009), 517–539. https://doi.org/10.1007/s11192-007-2046-8

[2] Giovanni Abramo, Ciriaco Andrea D'Angelo, and Francesco Rosati. 2016. Gender bias in academic recruitment. *Scientometrics* 106, 1 (2016), 119–141.

[3] Solon Barocas and Andrew D. Selbst. 2016. Big Data's Disparate Impact. *California Law Review* 104 (2016), 671. https://api.semanticscholar.org/CorpusID:143133374

[4] Brink, van den Brink Marieke, Brouns Margo, and Sietske Waslander. 2006. Does Excellence have a Gender? A National Research on Recruitment and Selection Procedures for Professorial Appointments in the Netherlands. *Employee Relations* 28 (11 2006), 523–539. https://doi.org/10.1108/01425450610704470

[5] Annunziata Calabrese, Maria Cristina Fede, Valeria Naciti, and Nicola Rappazzo. 2021. Female careers in Italian universities: The role of gender budgeting to achieve equality between women and men. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie/Cracow Review of Economics and Management* 5 (989) (2021), 31–47.

[6] Michelle I. Cardel, Emily Dhurandhar, Ceren Yarar-Fisher, Monica Foster, Bertha Hidalgo, Leslie A. McClure, Sherry Pagoto, Nathanial Brown, Dori Pekmezi, Noha Sharafeldin, Amanda L. Willig, and Christine Angelini. 2020. Turning Chutes into Ladders for Women Faculty: A Review and Roadmap for Equity in Academia. *Journal of Women's Health* 29, 5 (2020), 721–733. https://doi.org/10.1089/jwh.2019.8027 arXiv:https://doi.org/10.1089/jwh.2019.8027 PMID: 32043918.

[7] Magnus Carlsson, Henning Finseraas, Arnfinn H. Midtboen, and Guobjorg Linda Rafnsdottir. 2021. Gender Bias in Academic Recruitment? : Evidence from a Survey Experiment in the Nordic Region. *European Sociological Review* 37, 3 (2021), 399–410. https://doi.org/10.1093/esr/jcaa050

[8] Alexandra Chouldechova. 2017. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data* 5, 2 (2017), 153–163. https://doi.org/10.1089/big.2016.0047 PMID: 28632438.

[9] Sam Corbett-Davies, Johann D. Gaebler, Hamed Nilforoshan, Ravi Shroff, and Sharad Goel. 2024. The measure and mismeasure of fairness. *J. Mach. Learn. Res.* 24, 1, Article 312 (mar 2024), 117 pages.

[10] Jesse Davis and Mark Goadrich. 2006. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning* (Pittsburgh, Pennsylvania, USA) *(ICML '06)*. Association for Computing Machinery, New York, NY, USA, 233–240. https://doi.org/10.1145/1143844.1143874

[11] Maria De Paola, Michela Ponzo, and Vincenzo Scoppa. 2017. Gender differences in the propensity to apply for promotion: evidence from the Italian Scientific Qualification. *Oxford Economic Papers* 69, 4 (04 2017), 986–1009. https://doi.org/10.1093/oep/gpx023 arXiv:https://academic.oup.com/oep/article-pdf/69/4/986/20501975/gpx023.pdf

[12] Ministero dell'Istruzione dell'Università e della Ricerca. 2023. Cerca Università. http://cercauniversita.cineca.it/php5/docenti/cerca.php

[13] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (Cambridge, Massachusetts) *(ITCS '12)*. Association for Computing Machinery, New York, NY, USA, 214–226. https://doi.org/10.1145/2090236.2090255

[14] Giordano d'Aloisio, Andrea D'Angelo, Antinisca Di Marco, and Giovanni Stilo. 2023. Debiaser for Multiple Variables to enhance fairness in classification tasks. *Information Processing & Management* 60, 2 (2023), 103226. https://doi.org/10.1016/j.ipm.2022.103226

[15] Tom Fawcett. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27, 8 (2006), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010 ROC Analysis in Pattern Recognition.

[16] Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and Removing Disparate Impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Sydney, NSW, Australia) *(KDD '15)*. Association for Computing Machinery, New York, NY, USA, 259–268. https://doi.org/10.1145/2783258.2783311

[17] Marianna Filandri and Silvia Pasqua. 2021. 'Being good isn't good enough': gender discrimination in Italian academia. *Studies in Higher Education* 46, 8 (2021), 1533–1551. https://doi.org/10.1080/03075079.2019.1693990 arXiv:https://doi.org/10.1080/03075079.2019.1693990

[18] Mary Frank Fox. 2005. Gender, family characteristics, and publication productivity among scientists. *Social Studies of Science* 35, 1 (2005), 131–150.

[19] Batya Friedman and Helen Nissenbaum. 1996. Bias in computer systems. *ACM Trans. Inf. Syst.* 14, 3 (jul 1996), 330–347. https://doi.org/10.1145/230538.230561

[20] Camilla Gaiaschi and Rosy Musumeci. 2020. Just a Matter of Time? Women's Career Advancement in Neo-Liberal Academia. An Analysis of Recruitment Trends in Italian Universities. *Social Sciences* 9, 9 (2020). https://doi.org/10.3390/socsci9090163

[21] Krista Lynn Glass, Christy; Minnotte. 2010. Recruiting and hiring women in STEM fields. *Journal of Diversity in Higher Education vol. 3 iss. 4* 3 (2010). Issue 4. https://doi.org/10.1037/a0020581

[22] Moritz Hardt, Eric Price, Eric Price, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Eds.), Vol. 29. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf

[23] Levke Henningsen, Alice H. Eagly, and Klaus Jonas. 2022. Where are the women deans? The importance of gender bias and self-selection processes for the deanship ambition of female and male professors. *Journal of Applied Social Psychology* 52, 8 (2022), 602–622. https://doi.org/10.1111/jasp.12780 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jasp.12780

[24] Joanne Kenney, Susanna Ochoa, Mohamed Abdalhameed Alnor, Benneth Ben-Azu, Luciana Diaz-Cutraro, Royhaan Folarin, Avril Hutch, Hilmar Klaus Luckhoff, Cintia R. Prokopez, Nicole Rychagov, Bakare Surajudeen, Louise Walsh, Tonya Watts, and Elisabetta C. del Re. 2022. A Snapshot of Female Representation in Twelve Academic Psychiatry Institutions Around the World. *Psychiatry Research* 308 (2022), 114358. https://doi.org/10.1016/j.psychres.2021.114358

[25] Giulio Marini and Viviana Meschitti. 2018. The trench warfare of gender discrimination: evidence from academic promotions to full professor in Italy. *Scientometrics* 115, 2 (May 2018), 989–1006. https://doi.org/10.1007/s11192-018-2696-8

[26] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6, Article 115 (jul 2021), 35 pages. https://doi.org/10.1145/3457607

[27] Maria De Paola and Vincenzo Scoppa. 2011. *Gender Discrimination and Evaluators' Gender: Evidence from the Italian Academy.* Working Papers 201106. Università della Calabria, Dipartimento di Economia, Statistica e Finanza "Giovanni Anania" - DESF. https://ideas.repec.org/p/clb/wpaper/201106.html

[28] Ilenia Picardi. 2019. The Glass Door of Academia: Unveiling New Gendered Bias in Academic Recruitment. *Social Sciences* 8, 5 (2019). https://doi.org/10.3390/socsci8050160

[29] Jared Russell, Sheri Brock, and Mary Rudisill. 2019. Recognizing the Impact of Bias in Faculty Recruitment, Retention, and Advancement Processes. *Kinesiology Review* 8 (11 2019), 1–5. https://doi.org/10.1123/kr.2019-0043

[30] Denise Sekaquaptewa, Koji Takahashi, Janet Malley, Keith Herzog, and Sara Bliss. 2019. An evidence-based faculty recruitment workshop influences departmental hiring practice perceptions among university faculty. *Equality, Diversity and Inclusion: An International Journal* (2019).

[31] Seema S. Sonnad and Lisa M. Colletti. 2002. Issues in the recruitment and success of women in academic surgery. *Surgery* 132, 2 (2002), 415–419. https://doi.org/10.1067/msy.2002.127694

[32] Delys Todd, Patricia; Bird. 2000. Gender and promotion in Academia. *Equal Opportunities International 2000-dec vol. 19 iss. 8* 19 (dec 2000). Issue 8. https://doi.org/10.1108/02610150010786166

[33] Department of Economic United Nations and Social Affairs Sustainable Development. 2015. Transforming our world: the 2030 Agenda for Sustainable Development. , 16301 pages. https://sdgs.un.org/2030agenda

[34] Melanie E. Ward. 2001. Gender and Promotion in the Academic Profession. *Scottish Journal of Political Economy* 48, 3 (2001), 283–302. https://doi.org/10.1111/1467-9485.00199 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9485.00199

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009