KARYA ILHAM PUBLISHING
CONNECTING MINDS, BRIDGING IDEAS, INSPIRING READERS

Journal of Advanced Research in Computing and Applications

Journal homepage:
https://karyailham.com.my/index.php/arca
ISSN: 2462-1927

Predictive Modelling of the Gender Pay Gap using Machine Learning

Loh Siew Chin[1], Maizatul Akmar Ismail[1], Jamallah Mohammed Zawia[1,*]

[1] Department of Information System Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 29 July 2025
Received in revised form 22 August 2025
Accepted 31 August2025
Available online 4 September 2025

Keywords:
Gender pay gap; machine learning; predictive modelling; SHAP; labor economics

ABSTRACT

The gender inequity in pay remains a major issue to the world's economy and society. The complex dynamic non-linearity that drives wage dispersion is not well represented in traditional econometric methods. This paper accounts for this issue by employing state-of-the-art powerful machine learning methods to construct a predictive model of the GWE. We compare five prediction models: Linear Regression, Random Forest, XGBoost, Support Vector Regression (SVR), and Neural Network using data from the latest U.S. Census Bureau's American Community Survey for a span of 2017 to 2023. The goal was accurate evaluation of the Ratio, which represents the value of women median wages divided by men median wages. The performance of the model and its interpretability were strictly evaluated via 10-fold cross-validation and SHapley Additive exPlanations (SHAP), respectively. The results revealed that tree-based ensemble models considerably outperformed other methods, with Random Forest having the best performance (R = 0.994, MAPE = 0.443%), followed by XGBoost (R = 0.993, MAPE = 0.450%). Occupation was the most important source of the gender gap; this was found using both approaches. On the other hand, Linear Regression was relatively successful (R = 0.655) and Support Vector Regression failed (R = -0.303). These findings are a reaffirmation of the ability of machine learning, mainly Random Forest and XGBoost, as powerful and interpretable predictive tools. Such strategies provide policymakers with useful data for developing policies by occupation as well as targeted actions to minimize the gender pay gap and promote equity in the workplace.

1. Introduction

1.1 Research Background

Gender equality in the workplace remains a global priority, yet disparities in compensation between men and women continue to be a major concern. Despite improvements in education, labor force participation, and economic development, the gender pay gap persists across countries and sectors. This issue not only reflects inequities in the labor market but also signals broader systemic barriers that need deeper investigation and targeted policy responses.

The gender pay gap is the variation in median compensation between men and women in a broad range of jobs and fields. There are still significant differences in income, even though gender equality, education, and participation in the labor have all gotten better around the world. The International

* *Corresponding author.*
E-mail address: jamala.zawia@um.edu.my

https://doi.org/10.37934/arca.39.1.101110

Labor Organization (ILO) states that women around the world make roughly 20% less than men, even when you factor in things like education and experience. This persistent gap highlights that macroeconomic improvements alone are insufficient to eliminate wage disparities. Even when the economy is stronger, there are still big pay gaps between men and women in developed countries. This demonstrates that just making the economy better doesn't fix gender-based pay gaps. Discrimination, occupational segregation, and the fact that women don't have enough representation in high-paying jobs are some of the structural problems that lead to these wage gaps, as per Blau and Lawrence [1]. Moreover, it is essential that pay transparency laws are designed thoughtfully so as not to have unintended consequences, for example, creating a chilling effect, discouraging open discussions around pay as highlighted by Bishu and Mohamad [2].

The Oaxaca-Blinder decomposition and other traditional econometric methods have been used a lot to look at the gender wage gap. They break it down into explained (observable elements) and unexplained (possibly discriminatory) parts according to Oaxaca [3]. These linear econometric models, on the other hand, make the mistake of assuming simple correlations between variables and not taking into account the many interactions and non-linearities that are a part of how the labor market works. This constraint means that we need to use other analytical methods that can capture the several factors that affect wages and give us more information.

In the same context, England [4] has mentioned that women have an outsized presence in lower-paying industries from teachers to clerks to nurses while men dominate higher-paying ones, from engineering and technology to banking and finance, according to some researchers. And Dias [5] stated that occupational segregation is another economic factor. Low-paying industries like education, healthcare support, retail, and administration continue to overemploy women.

To sum up, the gender pay gap is still a serious issue across industries and countries, even as education, economic conditions, and workforce participation improve. Just making the economy stronger does not fix wage inequality. Tackling this problem means going beyond surface-level solutions and addressing deeper structural issues like discrimination, occupational segregation, and lack of representation in high-paying jobs. It also means using better analytical methods that reflect the real complexity of the labor market. Only then can we fully understand and begin to close the gender wage gap.

1.2 Literature Review

While traditional methods have been used to analyze these gaps, they often fail to uncover complex patterns and hidden biases as per Strittmatter and Conny [6], machine learning has significantly affected the domestic labor market with new techniques applied to the study of the socio-economic (e.g., wage disparities) phenomenon [7]. It can also be employed in labor economics to provide new ways of addressing wage differentials, with several benefits over the traditional econometric approach. In his work, Bonaccolto-Töpfer and Stephanie [9] mentioned that unlike traditional linear regression models, ML models can identify more complex, non-linear forms of relationships among variables, such as occupation, education, gender, tenure, and sector of industry, leading to a more advanced explanation of wage determination and differentials.

According to [15], machine learning methods may outperform traditional statistical techniques by unveiling hidden patterns, modeling nonlinear relationships, and making more precise predictions. In the study of wage gaps, machine learning can draw upon mass demographic, occupational, and sectoral data to make more precise and transparent generalizations about wage. For instance, supervised learning models are often applied to predict salaries based on variables like education, experience, job title, and location, while isolating gender as an independent factor as mentioned by

Penner *et al.,* [8]. Similarly, Bonaccolto-Töpferand Stephanie [9] has proved that machine learning technologies outperform traditional econometrics sufficiently well in capturing the complications of wage inequality. For example, Random Forests and G.B. models can be utilized to discover which are the most important factors for wage inequality (occupation, educational level, industry). It enables the execution of policy modifications to be pinpointed.

Nevertheless, the application of sophisticated machine learning methods raises concerns regarding the interpretability and fairness of the resulting models. SHapley Additive exPlanations (SHAP) is one such approach. This provides clear, feature-specific interpretations of machine learning model predictions that can be applied in policy-making [10]. Even though there is potential in these new techniques, some critical limitations persist in the methodologies. The studies of Binns *et al.,* [11] on the gender pay gap typically rely on primitive descriptive statistics or traditional regression methods and do not make the most of machine learnings power to predict and explain phenomena.

With the increasing availability of granular socioeconomic data and faster advances in machine learning algorithms, there is a pressing need to bridge this methodological gap. Specifically, applying ML to U.S. Census Bureau data from 2017 to 2023 offers an unprecedented chance not only to improve earnings gap forecasting but also to uncover dynamic patterns over time, industry, and place [12]. This approach can offer insightful suggestions to policymakers, businesses, and advocacy groups striving for the eradication of the gender pay gap.

In addition, machine learning approaches have been widely accepted successfully for predictive analytics but are not fully understood when applied to the social and economic domain. This runs like a red line through the lines of reasoning about the significance of robust and interpretable algorithms [14].

The originality of our research consists in covering this gap using a large occupational sample drawn from the U.S. Census Bureau's American Community Survey (ACS 2017-2023) and comparing the performance of different predictive models, namely, Linear Regression, Random Forest, XGBoost, Support Vector Regression (SVR), and Neural Networks, in predicting and explaining the gender pay gap. The primary goals are the growth of accurate predictive models, identification of the most important factors affecting wage gaps, and the emphasis on the trend over time. Our findings arm policymakers with an evidence-based toolkit for designing targeted interventions to tackle gender-specific wage discrepancies and foster equitable workplaces.

2. Methodology

2.1 Data Collection

The study is based on publicly accessible data from the U.S. Census Bureau's American Community Survey (ACS) for 2017-2023. The ACS provides rich information about labour market outcomes, including occupation-specific median earnings by gender. The dataset has essential features: Occupation, Median Earnings of Men, Median Earnings of Women, Earnings Ratio (the ratio of women's median earnings to men's), and Year. Data quality was also verified meticulously and systematically, those with inconsistencies and missing values being discarded to guarantee authenticity and credibility.

2.2 Target Variable

The primary outcome variable of importance was the Earnings Ratio, alternatively described in the Census Bureau as the percent of the median earnings of women relative to men in the same occupation during the same year. An Earnings Ratio of 100 indicates complete wage equality, and values lower than 100 indicate levels of inequality.

2.3 Feature Engineering and Selection

The issue of gender pay disparities is particularly pronounced in STEM fields (Science, Technology, Engineering, and Mathematics). Engineering, for example, remains a male-dominated domain where women face significant challenges in both compensation and career progression, particularly in subfields such as mechanical and electrical engineering. Similarly, the technology sector frequently reports salary gaps in critical roles like software engineering, data science, and IT management, despite these areas being heralded as meritocratic. According to Edelsztein and Sebastián [13], these disparities highlight the need for more nuanced analytical approaches to identify root causes and propose actionable solutions.

From the ACS dataset, we first derived features such as Occupation, Median Earnings of Men, Median Earnings of Women, and Year. The model's performance was improved by adding further engineered features, such as:

i. Wage Gap (US$): Absolute payment ensures the difference between men and women.
ii. Ratio of Wage Gap: Wage gap relative to men's median earnings to account for the wage gap in different earnings levels.
iii. Sectoral Aggregations: Groups of occupations (e.g., STEM, Healthcare, Education, Management) based on SOC codes.
iv. Logarithmic Transformations: Earnings were log-transformed to deal with right-skewness and to stabilize variance.

Variable selection was directed by correlation and domain knowledge and further refined with repeated assessment to prevent multicollinearity and to maintain addition of relevant predictors.

2.4 Exploratory Data Analysis (EDA)

Initial analysis consisted of examining the distributions of earnings by gender and occupation over the study period. The trends show no indication of diminishing gender differences, which are the widest in high-paying fields such as Finance, Engineering, and Technology. EDA revealed the significance of models that can capture non-linear links between the predictors and outcomes.

2.5 Machine Learning Models

Five machine learning models were shortlisted after screening their predicting accuracy and interpretability:

i. Linear Regression: Baseline model that models linear relationships in features.
ii. Random Forest: Ensemble method based on trees suitable for handling complex, non-linear relations.
iii. XGBoost: Gradient boost model which has strong predictive power and remarkable sensitivity to the temporal dynamics.
iv. Support Vector Regression (SVR): Considered due to its ability to model the boundary relationships, although not as effective with socio-economic data.

v. Neural Networks: Investigated for their power to capture complex relations, interpretability was a bottleneck.

2.6 Model Validation

In order to test the robustness, and to estimate and compare the prediction accuracy of the models, 10-fold cross-validation was used: the raw dataset was divided into ten subsets and the process of model training and testing was done iteratively.
The performance of our model was evaluated on following metrics:

i. R-squared (R^2): Explained variance.
ii. RMSE (Root Mean Squared Error): Measure of the accuracy of a predictor.
iii. MAPE: Relative prediction accuracy.

In addition, there was a temporal hold-out validation (training on prior years, testing on subsequent years) to assess the robustness and predictive generalizability of the models over time.

2.7 Model Interpretability

Interpretability was paramount for actionable policy recommendations. The input features importance in the final model predictions was revealed by SHapley Additive exPlanations (SHAP) values estimation.
SHAP helped in pinpointing key factors and interactions contributing to the gender wage gap, offering an actionable framework to policymakers for generating informed policy interventions.
This rigorous methodological procedure guarantees that machine learning's predictive and interpretative power is utilized to provide strong, precise, and policy-relevant findings about the gender pay gap.

3. Results
3.1 Descriptive Statistics of Wage Differences

Analysis of earnings from 2017-2023, as shown in Figure 1, is a continuous, and in some cases worsening, gender pay gap among multiple occupations. The Earnings Ratio - the ratio of women's to men's median earnings - fell well below parity, reflecting persistent gender wage disparities. Men's median earnings averaged \$40,152, which was significantly higher than that of women (\$3,392), highlighting increased occupational and sectorial disparities. Though there had been some progress separating real wage parity over the decades, meaningful wage gaps were not addressed, notably in high paying industries such as finance, technology, and engineering. Figure 2 depicts the Distribution of Women's to Men's Earning Ratio (2023).

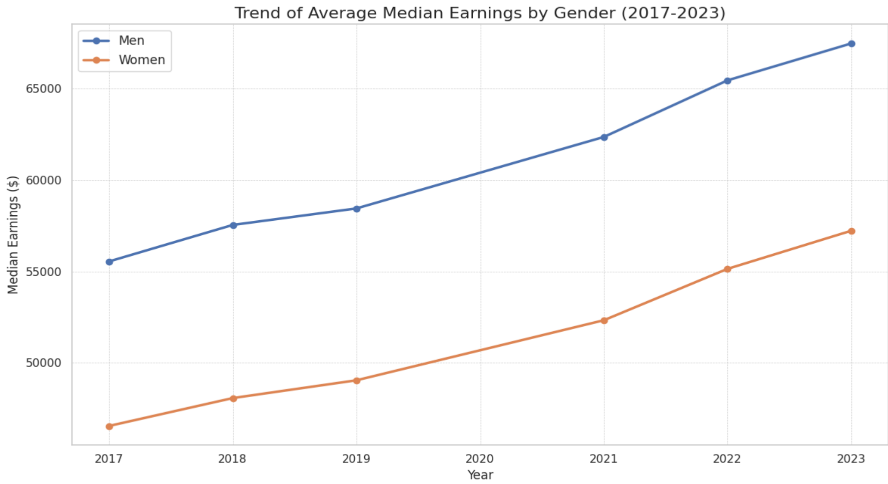

Fig. 1. Trend of average median earnings by gender (2017-2023)

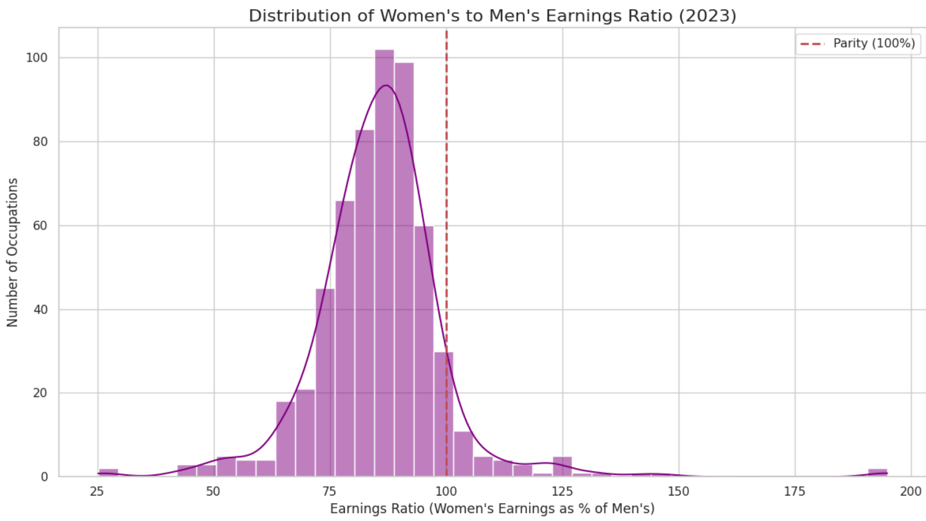

Fig. 2. Distribution of women's to men's earning ratio (2023)

3.2 Comparison of Predictive Models

The 10-fold cross-validation was used to compare the performance of five machine learning models, which also are explained in Table 1:

i. Random Forest turned to be the best model with R^2 value 0.994 and MAPE of 0.443%.
ii. It was closely followed with XGBoost (R^2=0.993, MAPE=0.450%), highlight its capability for modelling subtle temporal phenomena.
iii. Linear Regression, an interpretable model, provided a fine fit (R^2=0.655 indicating that the model has little capacity to model complex, non-linear relationships).
iv. Neural Networks showed poor explanation capability (R^2=0.023), possibly resulting from a lacking data complexity or model setup.
v. SVR was inappropriate as well, as it produced negative R^2 scores (-0.303), which means that it was fundamentally not compatible to the numerical socioeconomic properties of the dataset.

Table 1
Model Validation Results (10-Fold Cross-Validation)

Model	RMSE (Mean)	R^2 (Mean)	MAPE (Mean)	Key Interpretation
Linear Regression	21.933	0.655	2.428%	Moderate explanatory power, struggles with non-linear wage gap patterns
Random Forest	2.964	0.994	0.443%	Best overall performance, near-perfect predictions
XGBoost	3.115	0.993	0.450%	Comparable to Random Forest, slightly better at temporal trends
SVR	42.716	-0.303	8.661%	Failed to generalize, unsuitable for this task
Neural Network	38.002*	0.023*	7.892%*	Underperformed due to data limitations

3.3 Importance of Feature and SHAP Analysis

SHapley Additive exPlanations (SHAP) offers a consistent interpretation for explaining complex models based on the contribution of each feature to any prediction [15]. For better interpretation, the SHAP was used to value which unveil key features contributing to wage imbalances as follows:

i. Occupation was always among the most powerful predictors in all the models, which reflects the centrality of occupational segregation for understanding the wage inequality.
ii. Another factor strongly contributing to wage discrimination was Median Earnings of Male and Sectoral Groupings, where there is a bigger difference in wages in male-dominated fields that may be perceived as high-paying jobs, for instance, STEM or management sectors.
iii. The temporal analysis using SHAP values revealed a slow but irregular convergence of the wage gap between 2017 and 2023, illustrating the trends in different sectors.

The following Figure 3 shows a summary plot for the SHAP Analysis:

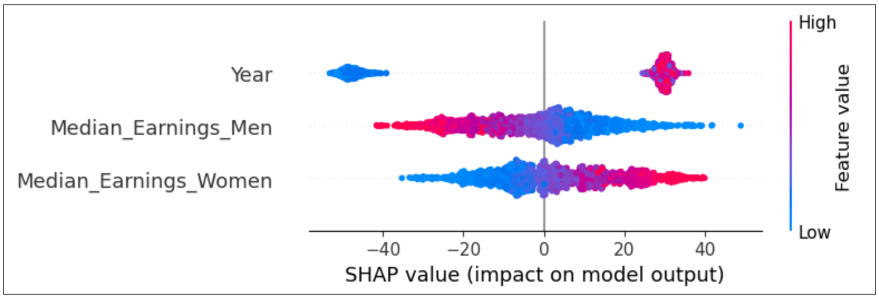

Fig. 3. SHAP summary plot

3.4 Wage Gap by Sector and Time Trends

Sector-specific analyses further illustrated disparities:

i. Females in Finance, Engineering, and Science had a markedly higher gender wage gap, where the Female Earnings Ratio was far from parity despite median wages being relatively high.
ii. Education and Healthcare, by comparison, had much smaller, though still significant, wage gaps, suggesting that overall, even in fields women dominate, disparities between the sexes are rife.

iii. Temporally, the wage gap was marginally reduced during years 2017–2023, although at various paces of reduction among different sectors.

3.5 Summary of Findings

The results indicate that the sophisticated machine learning techniques, that is, Random Forest and XGBoost, are able to detect complex nonlinear relationships with respect to gender wage discrimination. The models have improved predictive ability and are more interpretable as they clearly point out that occupational and sectoral dynamics are critical determinants of the gender pay gap. This study's results provide a rich, empirically sound base for policymakers interested in the development of focused, occupation-specific interventions to address workplace equity.

The following Figure 4 shows our work dashboard for the Gender Pay Gap Predictive Analysis:

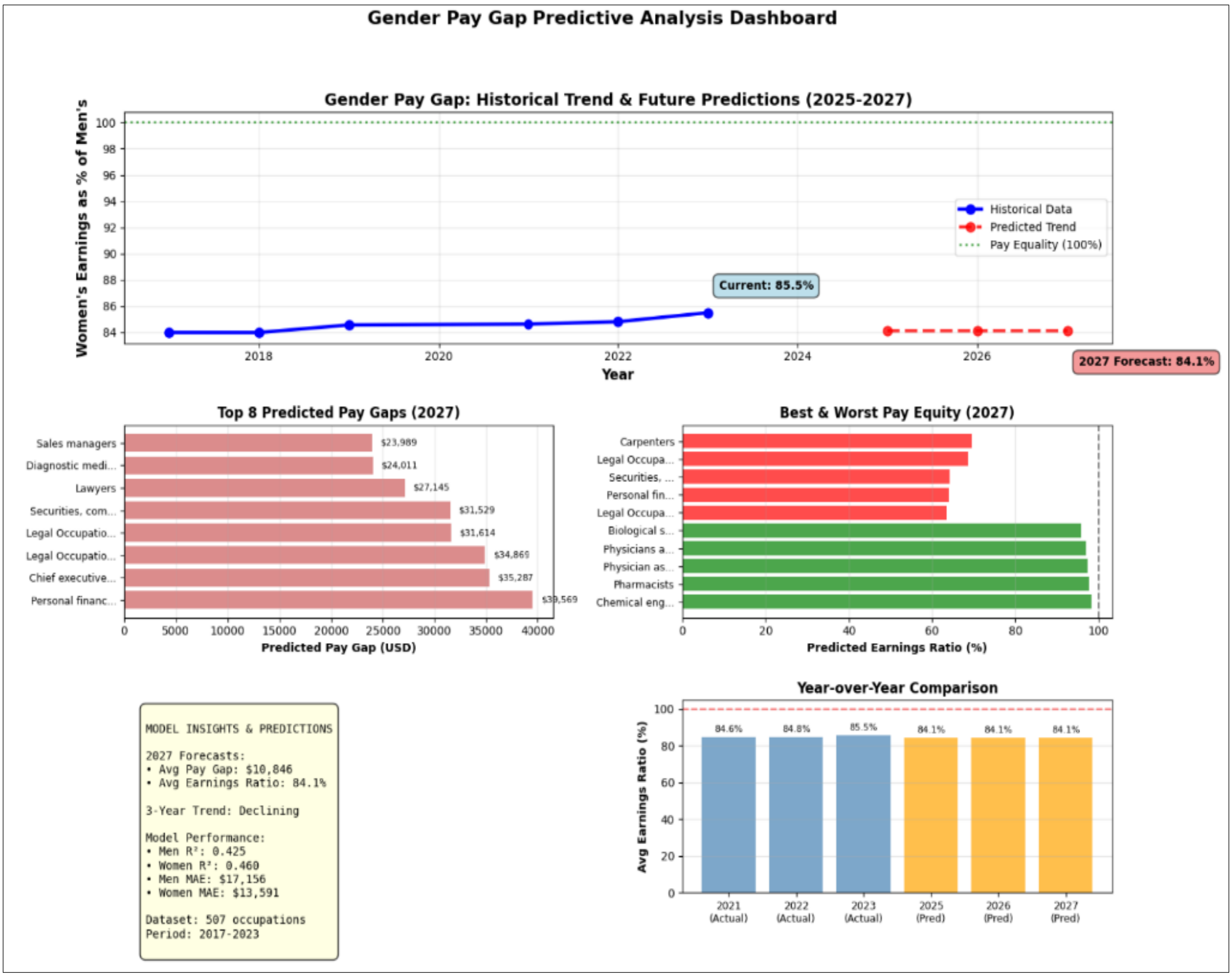

Fig. 4. Gender pay gap predictive analysis dashboard

4. Conclusions

In this study, we applied advanced artificial intelligence methods to analyze and predict gender wage gaps, using massive open-source occupational income data from the U.S. Census Bureau's American Community Survey (ACS, 2017-2023). The results reveal that traditional linear econometric models do not sufficiently capture the complex non-linear dynamics of the wage disparities. Several machine learning algorithms, including Random Forest, resulted in excellent models with high accuracy (R = 0.99) and low prediction error (MAPE < 0.5%). Our approach effectively captured intricate relationships among occupational groups, income strata, and workforce exposure time associated with wage gaps. Occupation was the primary contributor to the gender pay gap in all models, suggesting the presence of systemic occupational segregation. Highly paid industries like finance, tech, and engineering show some of the most startling gender discrepancies, suggesting enduring structural impediments faced by women in these fields. SHAP analysis enhanced interpretability by highlighting interacting factors and threshold effects that are useful for formulating focused policy interventions. This analysis highlights interpretable ML as an analytical utility for policymakers, promoting occupation-specific interventions to reduce gender wage gaps. Given the limited research in this area, we suggest that future research should expand the temporal scope and add more socio-demographic factors to understand the intricacies of wage gaps at their intersectional levels in order to inform evidence-based policies aiming at workplace equity.

References

[1] Blau, Francine D., and Lawrence M. Kahn. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55, no. 3 (2017): 789-865. https://doi.org/10.1257/jel.20160995
[2] Bishu, Sebawit G., and Mohamad G. Alkadry. "A systematic review of the gender pay gap and factors that predict it." *Administration & Society* 49, no. 1 (2017): 65-104. https://doi.org/10.1177/0095399716636928
[3] Oaxaca, Ronald. "Male-female wage differentials in urban labor markets." *International economic review* (1973): 693-709. http://dx.doi.org/10.2307/2525981
[4] England, Paula. "The gender revolution: Uneven and stalled." *Gender & society* 24, no. 2 (2010): 149-166. https://doi.org/10.1177/0891243210361475
[5] Dias Gonçalves, Mariana. "Gender Disparities in Occupational Outcomes: A Global Overview." *Journal of Labor Studies* 45, no. 2 (2024): 125–147.
[6] Strittmatter, Anthony, and Conny Wunsch. "The gender pay gap revisited with big data: do methodological choices matter?." arXiv preprint arXiv:2102.09207 (2021). https://doi.org/10.48550/arXiv.2102.09207
[7] Athey, Susan. "The impact of machine learning on economics." In *The economics of artificial intelligence: An agenda*, pp. 507-547. University of Chicago Press, 2018. https://doi.org/10.7208/chicago/9780226613475.001.0001
[8] Penner, Andrew M., Trond Petersen, Are Skeie Hermansen, Anthony Rainey, István Boza, Marta M. Elvira, Olivier Godechot et al. "Within-job gender pay inequality in 15 countries." *Nature human behaviour* 7, no. 2 (2023): 184-189. https://doi.org/10.1038/s41562-023-01523-x
[9] Bonaccolto-Töpfer, Marina, and Stephanie Briel. "The gender pay gap revisited: Does machine learning offer new insights?." *Labour Economics* 78 (2022): 102223. https://doi.org/10.1016/j.labeco.2022.102223
[10] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "" Why should i trust you?" Explaining the predictions of any classifier." In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135-1144. 2016. https://doi.org/10.1145/2939672.2939778
[11] Binns, Reuben, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. "'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions." In *Proceedings of the 2018 Chi conference on human factors in computing systems*, pp. 1-14. 2018. https://doi.org/10.1145/3173574.3173951
[12] Varian, Hal R. "Big data: New tricks for econometrics." *Journal of economic perspectives* 28, no. 2 (2014): 3-28. https://doi.org/10.1257/jep.28.2.3
[13] Edelsztein, Valeria Carolina, and Sebastián Ariel Waisbrot. "Breaking down the Gender Pay Gap through a machine learning model." Convergencia 30 (2023). https://doi.org/10.29101/crcs.v30i0.20656

[14] Doshi-Velez, Finale, and Been Kim. "Towards a rigorous science of interpretable machine learning." *arXiv preprint arXiv:1702.08608* (2017). https://doi.org/10.48550/arXiv.1702.08608

[15] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." *Advances in neural information processing systems* 30 (2017). https://doi.org/10.48550/arXiv.1705.07874