9th International Conference on Prestigious Applications of Intelligent Systems – PAIS@ECAI2020
Santiago de Compostela, Spain

Equal Pay for Equal Competences : A Statistical Approach to Address Equal Pay Gap

Andra Anoaica and **Ahmed Ben Hassine** and **Léa A. Deleris** [1]

Abstract.

Gender wage equality is something that a majority of modern companies work towards. Measuring and addressing gender pay gap in practice can however be challenging due to the variety of roles, educational paths, experiences and responsibilities that compose any given company workforce. In this paper we propose a methodology to (i) determine the level of pay inequality between men and women within a single organisation and to (ii) suggest adjustments to address it. Our approach is novel in the sense that previous studies were done on homogeneous cherry-picked populations from different enterprises at a country level. By contrast, our model aims at decision support for compensation exercises within a single organization. The methodology is applied in two steps. First we evaluate the health of the overall organisation from the gender gap point of view. We combine expert-driven rules with unsupervised learning methods to identify individuals that are comparable, and then provide metrics to assess the level of gender-based pay gap within these groups. For cases where such a gap is ascertained, we build a predictive model for the gender neutral salary of each individual based on comparison with his/her closest peers within the workforce. Once the individuals that have a deficit in terms of compensation are identified, we test how the budgetary constraints reflect on a catch up plan. We have applied this methodology to a European organisation of about 1500 employees to support the HR department in its efforts to assure competitive compensation. The results confirm the existence of differences in compensation between men and women but also that the adjustments proposed by our models do lead to narrowing the pay gap.

1 INTRODUCTION

The motivation for eliminating the pay gap between men and women can be very diverse, from a self-evident ethical one, to a more market-oriented supported by neoclassical economic models which see a high correlation between gender equality and enterprise competitiveness. The equal pay problem can be summarised as follows: Although men and women might occupy similar positions, it can happen that they do not receive equal compensation. This can hurt the competitiveness of employees compensation.

A recent report on gender inequality released by UN Women, the HeForShe IMPACT Report [10], which tackles a wide variety of gender based discrimination areas, showed statistics that confirm this economic hypothesis. According to OECD, the cost of gender-based discrimination in social institutions is $6 trillion which represent 7.5% of the global GDP. Ensuring an equal footing for women and men from an economic standpoint is good for business.

This paper proposes a novel methodology for measuring and correcting the gender gap in test phase today within one of the biggest financial institutions in France. As opposed to previous studies the population of the study is diverse with more than 250 different job titles. Our approach is based on a combination of unsupervised machine learning and statistical techniques that are easy to adapt when building a catch up plan. This alternative to the method typically used that consists in looking at parameters within a regression model to quantify for the disparity is more intuitive for a manager that wants to implement a catch up plan. Moreover, once we obtain the estimation of the gender gap, we propose a method for wealth distribution under different budgetary constraints which responds to real-life resource constraints.

Our research has looked at gender, but could also tackle other types of discrimination.

The remainder of the paper is organized as follows. First, we present a review of the associated research work in Section 2. Section 3 and Section 4 then provide details regarding respectively the data and the model used in our approach. Section 5 then describes the results from the application of the approach to a real world company, highlighting benefits and limitations.

2 RELATED WORK

Gender pay inequality is a vibrant research domain, however, most exclusively through the prism of social science. Indeed, to the best of our knowledge, the literature on gender disparities in the labor market examines whether an unexplained pay disparity remains after controlling for individual characteristics that are expected to influence earnings, with control variables serving as proxies for productivity. The studies are diverse in terms of settings, be it the perimeter of the study (industry level, country level) or the method of data collection (surveys, company database). The number of quantitative methods that measure and detect the disparities is quite limited and it is better adapted to sociological studies than to assist an HR department.

In [7], the authors take the two main sources of earnings in the UK, the Annual Survey of Hours and Earnings (ASHE), and the Labour Force Survey (LFS) and compute the gender pay gap mainly using the difference in percentage between men and women when taking into consideration different filter factors such as age, occupation, region, distribution of hours, company size, family characteristics, educational attainment. The authors show that a pay gap still exists for all the categories combinations enumerated above.

Other authors [11] rely on more sophisticated methods to quantify the gender gap. A few different IT occupations in the United States are included in the study: computer system analysts, software

[1] BNP PARIBAS, France, email: {andra.anoaica, ahmed.benhassine, lea.deleris}@bnpparibas.com

engineers, information systems scientists. The authors use multiple linear regressions and propensity score analysis to estimate, for each occupation, the overall gender salary gaps. Both the propensity score and linear regression methods can be used to make descriptive comparisons of the salaries of similar men and women. This paper also proposes a novel methodology to handle data from surveys in this type of settings as some population might be over represented by the sample in relation to the actual population.

Another stream of research on gender pay gap looks more broadly at the problem from the perspective of choices with respect to family and household. For instance, [5] shows that even when taking into account these factors, discrimination was still present at the end of the 20th century. A more recent study from [2] attempts to eliminate possible variability of factors related to household choices and focus on the job characteristics. They apply a simple Oaxaca-Blinder [6] decomposition on a data set containing high-level executives in US corporations. They demonstrate that in this setting when controlling for measurable characteristics, for women and men who operate in corporations of similar size with a comparable profile, there is no significant gender gap. [9] is also a study relying on the Oaxaca-Blinder decomposition to highlight the gender gap, this time comparing the results between multiple countries. The authors find that the gender employment gaps are important in understanding the cross-country differences. The non-random selection into work is controlled by a wage imputation for the data points that are marked as unemployed.

In [4] the authors study the pay gap within one large private firm in the US. The current and starting salaries are the endogenous variables that are modeled using least square based on human capital variables, such as age, education dummies, job titles, job performance as explanatory features. The study highlights that an important factor accounting for the wage gap is the entry salary, implying that current salary imbalances are partially the result of a one time salary shortfall.

However, none of the above studies aim at designing a catch up plan for bridging the salary gap between men and women. The methods presented are better suited for assessing the gap without acting upon it. More than just a sociological study of discrepancies, we propose a pragmatic approach to estimate and address the gender gap at the level of an organisation. We evaluate the general health of an organisation from an equal pay point of view, compute the payment gap for every individual and propose a method for bridging the gap under budgetary restrictions.

3 DATA

The proposed methodology was tested on 1505 employees within the same organisation, but with different job types and responsibility levels. All employees work in the same country and are subject to the same labor law context and economic conditions. The data was preprocessed through minor transformations in order to preserve the privacy of employees. Specifically, the names of the employees were encrypted so the data analysts could work on data involving their colleagues without being able to precisely identify them. Furthermore, salaries were linearly transformed to preserve distributional properties while being meaningless to the researchers (as if it were an unknown currency).

Our data is separated into two sources: (i) Internal information from the Human Resource department on each employee (ii) External benchmark classification data. The list of features contained in the Human Resources (HR) file is provided in Table 1.

Variable Name	Details
Gender	54%Women, 46%Men
Age	Median age: 43 years old
Experience within the organisation	Median Experience: 14 years
Management position	21% managers
Job title	200+ job titles
Sub-department name	Number of sub-departments: 3
Team name	Number of Teams: 45
Salary in the Current year	Median Salary: 165 m.u.

Table 1. Data Description

In addition to the information from the HR department, each employee is mapped to an external benchmark taxonomy, which is normally used for external compensation benchmarks. This exercise is carried out annually (and in fact manually by the HR department team) to ensure the competitiveness of employees compensation with respect to peers working with competitors. While we cannot provide specific details for intellectual property concern, we illustrate with a simple example : A software developer would be mapped for instance to "14.Product Development" family of roles and further to "14.03 Software Development". Then, that person would have an indicator for whether he/she is a manager or an expert and his/her associated level of expertise between 1 and 4. Finally, starting from the CEO of the organisation, each person is assigned an overall responsibility level between 0 and 30, with 30 corresponding notionally to the responsibilities of the CEO of a Fortune 500 company. So in the end, a person is mapped to five data item (i) job family (14), (ii) job role (14.03) (iii) manager indicator (iv) expertise level and (v) overall responsibility (e.g., 14).

4 METHODOLOGY

In this section, we present in further details the methodology that we have developed. Beyond wage gap, which directly compares median salary earned by men to median salary earned by women, we are interested in *equal pay*. To quantify equal pay, we need to consider the nature of the job performed (role and responsibilities).

We propose a first step to evaluate the overall health/status of compensation between men and women which is based on unsupervised methods and expert rules. The underlying challenge is to find employees whose characteristics affecting productivity on the job and impact within the organisation are as similar as possible. After achieving relatively homogeneous clusters of employees, we can compare salaries using a series of statistical methods suitable for assessing inequality within and between populations. Once we have a picture of the company, the second step is to look at how much each individual should be paid relative to his or her peers. We rely on a k-NN (k Nearest Neighbours) as a supervised techniques to identify the subset of individuals to which a person resembles from a job type and responsibility perspective. We further look at the salary of this group in regards to our target employee and compute the difference between the mean of the group's salaries and the target employees salary. The last and final step is to see how to bridge the pay gap, i.e., how to allocate a salary increase budget optimally in the likely case where budget constraints imply that there are not enough resources to cover all the required adjustments.

4.1 Organisation Level Compensation Fairness

In this section we propose a method to quantify compensation fairness within the organisation at an aggregated level. As we are inter-

ested in distinguishing well among the various job roles and competencies, we start by clustering employees into homogeneous groups using a two-level clustering. The first level of clustering relies on HR experts and we refine the larger groups further through unsupervised methods. One challenge related to building homogeneous groups is to ensure that they are small enough to account for the variability among groups, but large enough so the statistical methods applied within groups are robust.

4.1.1 Two Level Clustering

The first level of clustering was proposed by the HR experts and is thus rule-based. They group job disciplines into six broader categories that are similar in terms of complexity of the tasks and contribution to the organisation performance. This way, each individual that has a mapping for the external benchmark is automatically mapped to one of the six job categories. While we cannot give the exact name of the six broad job categories for confidentiality reasons, we briefly describe them. The first job category refers to jobs that require strong analytical skills as well as some programming skills. The second category refers to jobs that are support functions for the organisation main business. The third category refers to jobs that are dealing law and regulations, The fourth and fifth category are strongly related to the organisation's business and are differentiated by the type of clients served. The last category is focused on the quality of the activities within the organisation.

For the second level of clustering, we rely on unsupervised learning techniques. This family of models is suitable for our use-case as it helps in finding previously unknown patterns in a data set without pre-existing labels. The features taken into consideration are: the age (as a proxy for overall experience) represented by age intervals, the study level, the career path type (managerial or non-managerial), the department to which the employee belongs.

Continuous variables such as age are discretized into equal intervals. As all the variables are ultimately categorical, they are represented as one hot encoding. We apply K-means clustering with Euclidean distance. This method is suitable for our use-case as we have transformed all our features into numerical representations. We prefer this clustering algorithm to others because it produces clusters relatively comparable in terms of size and guarantees convergence.

4.1.2 Statistical Indicators of Salary Fairness

Up until this point we have not taken into consideration the "Gender" variable in our analysis. The reason is that we want to construct clusters that ignore the potential discriminatory factor. It is only after we obtain the clusters that we will separate our population into two: men and women.

We measure the inequalities among the clusters as well as within each cluster. The measurements used are (i) Gini index for intra-cluster evaluation and (ii) Almost Stochastic Dominance as well as (iii) the difference in terms of mean and median salary for inter-cluster comparisons.

Gini index: is a statistical measure that accounts for the distribution of a variable (salary, income, wealth) within a population [8]. In other words, it measures the level of inequality of the distribution of a variable within the population. In our case we will apply the Gini index within the men population and women population of each cluster, to account for the inequalities that might occur outside the gender driven ones.

Almost Stochastic Dominance (ASD): First-Degree (also called First-Order) Stochastic dominance is a form of stochastic ordering (i.e, enabling to partially order probability distributions). In decision theory, this concept is often used to characterize the relationship between two lotteries (i.e., distributions over possible outcomes). Specifically, if two lotteries can be ordered via first-order stochastic dominance, then any expected utility maximizer decision maker will choose the dominating lottery, thus enabling to determine the choice of the decision maker with minimal knowledge about his/her utility function.

However, Stochastic dominance is a condition which is hardly seen in real-world settings. In [1] the authors experiment on deviations from the stochastic ordering that are small enough to conclude the dominance of one distribution over the other. This family of models include Approximate Stochastic Dominance or Almost Stochastic Dominance (ASD). They represent a relaxation of stochastic dominance, which allows small violations of stochastic dominance rules to avoid situations where most decision makers prefer one alternative to another, but stochastic dominance cannot rank them.

Almost stochastic dominance can be adapted to a variety of settings outside the decision theory domain where we need to compare between two probability distributions. In [3] the authors use ASD to compare two neural networks' empirical distribution scores on unseen data. This method is proved to be more robust than comparing single evaluations scores.

We include ASD in our methodology as we consider it to be a meaningful way to characterize the gap between the empirical distribution of men's salaries and the empirical distribution of women's salaries.

4.2 Individual Compensation Comparison

Beyond the understanding of compensation fairness at the organisation level, we are interested in determining for each woman if there are significant differences in compensation in comparison to her masculine peers. If it is the case, the next step is to see which amount would enable to attain a competitive level. Note that we focus here on adjusting women's salaries as they are typically earning less than men. However, we could do a symmetric exercise if it was rather men that were discriminated against.

Specifically, for each individual within each of the six categories, we identify the employees that are closest to him or her in the same category and rely on this subgroup to determine how much he or she should be paid. Specifically, we make use of k-nearest neighbors (k-NN) algorithm which assigns the average salary of the k-nearest neighbours as the predicted salary. The optimal number of neighbors K is determined using cross-validation and mean absolute error (MAE) on salary as a quality metric. This method is preferred to others due to its strong explainability power as for any given individual we can explain his or her salary by looking at the features of his her closest neighbors. This gives the liberty to the HR department to make their own judgements and not have a black box model that delivers a predicted salary without further explanations.

Because we are interested in determining how a woman would be paid if we did not know the gender, the algorithm is trained using cross-validation on the men population only in order to learn how specific features like age and responsibility level reflect on men's salaries. Once we create our prediction model, we apply it on women without taking into consideration the gender feature at any moment. This prediction enables us to take the women employee characteristics and transpose them to the men's prediction space. Thus, we

	Number of Men	Number of Women	Number of Employees	Median Age	Median Experience	Median Salary
Job cat. 1	69	42	111	35	3	179
Job cat. 2	74	106	180	42	12	179
Job cat. 3	170	207	377	43	12	184
Job cat. 4	249	341	590	43	18	141
Job cat. 5	66	19	85	38	10	219
Job cat. 6	73	89	162	46	17	185

Table 2. Level 1 Groups

have a blind evaluation of the salary through our model. We train one model for each of the six job categories obtained from the two level clustering. The exercise could have been carried the other way around too, with the model being trained on the women population and applied on the men. However, as we suspect that this approach will lower salaries and in real life, HR department are reluctant (if not prevented by law) to lower salaries, we prefer to use the salaries of the men's population as the reference.

In addition to the variables used in the clustering model presented in the previous section, we use additional variables so as to be able to capture finer nuances. Namely, we also consider the overall responsibility grade from the external benchmarks, an indicator for highly important managerial positions and talents within the organisation (i.e, high potential junior people). In addition, the age variable in used as a continuous variables as opposed to the age intervals used in clustering.

5 EXPERIMENTAL RESULTS

5.1 Organisation Level Perspective

In the first phase, we look at the mapping to the first level of clusters based only on the expert rules provided by the HR department. Before applying the algorithms we checked the integrity of our data and eliminated data points without compensation information. In addition, prior to the mapping, we undertook a partial data quality check by organising external benchmark validation sessions with a subset of the senior managers of the organization. This enabled to identify employees whose associated external benchmark discipline had to be updated but more importantly it enabled to involve management in the process, thus increasing their understanding of the process and ultimately the credibility of the exercise.

Table 2 presents for each of the six job categories the number of disciplines included and the number of employees that fall into each category. The groups are rather balanced regarding gender ratio, with the exception of the 5th job category. The number of employees within each cluster exceeds 100, again with the exception of job category 5.

We are interested in further refining these groups and thus follow with the second level of clustering relying on unsupervised learning. The algorithm generates 12 clusters : two for job categories 1 and 6 and three for job categories 3 and 4, while job categories 2 and 5 are not be further broken down into smaller classes as the number of employees within is already sufficiently small.

Table 3 provides details about the output clusters, indicating the cluster name, the Gini index for men and for women, the difference in percentage between the mean salary of men and women, the difference in percentage between the median salary of men and women and whether there is Almost Stochastic Dominance between men and women. In that case, we also report the value of the parameter ϵ which measures the distance between the two salary distributions. The closer ϵ is to 1, the closer to pure stochastic dominance for the two distributions.

Cluster	Gini men	Gini women	Δ mean	Δ median	ASD	ϵ
Job cat. 1.1	0.22	0.14	25%	21%	**YES**	0.99
Job cat. 1.2	0.14	0.16	1%	8%	NO	
Job cat. 2.1	0.15	0.16	11%	10%	**YES**	0.79
Job cat. 3.1	0.20	0.23	25%	20%	**YES**	1
Job cat. 3.2	**0.23**	**0.18**	29%	14%	**YES**	1
Job cat. 3.3	0.19	0.16	18%	14%	**YES**	1
Job cat. 4.1	0.18	0.14	18%	3%	**YES**	1
Job cat. 4.2	0.16	0.14	5%	0%	**YES**	1
Job cat. 4.3	**0.19**	**0.11**	31%	25%	**YES**	0.99
Job cat. 5.1	**0.19**	**0.13**	13%	4%	**YES**	0.99
Job cat. 6.1	0.16	0.18	2%	6%	**YES**	0.62
Job cat. 6.2	0.2	0.23	12%	21%	**YES**	0.71
Average	0.18	0.16	14%	12%		0.79

Table 3. Cluster Gap Metrics

The results show that there is a clear dominance of the men empirical salary distributions over the women empirical salary distributions for all the clusters with the exception of job category 1.2. In this particular case, even though almost stochastic dominance does not hold, the mean and median are higher for men, suggesting that the absence of almost stochastic dominance is due highly paid females in the top percentiles, while preserving important inequalities for the lower percentiles of the distribution. What is more striking is that in 8 out of 11 cases we have stochastic dominance, meaning that the distribution of men's salaries is superior to the distribution's of women's salaries. Interestingly enough, the Gini index provides additional insights: In some clusters (Job cat. 3.1, Job cat. 4.3, Job cat 5.1) there are high inequalities within the men's groups, while it is not the case for women. These discrepancies are reflected in the differences between the delta of the mean of the salaries and the delta of the median of the salaries, with the former being greater than the latter.

5.2 Individual Level Perspective

The individual salary prediction models based on k-NN are trained on the male employees from each cluster, using a 5-fold cross-validation in order to determine the optimal number of k closest neighbors. The training size for each cluster is the number of men reported in Table 2.

Table 4 presents the optimal number of neighbors (k), the mean absolute error (MAE), the mean absolute percentage error (MAPE), metrics for training set (i.e., male empliyees). It also indicates the number of men for which the algorithm predicted a higher salary than their actual one (Nb. M+), or lower than their actual salary (Nb. M-), and the difference between those two numbers (Δ Nb. = Nb. M+ - Nb. M-). We report the MAE and MAPE only on the men population because they capture the model performance, while if our hypothesis of discrimination between men and women is correct, than the MAE and MAPE results for women include not only the model performance but also a penalty due to discrimination. We can see that the optimal number of neighbors (k) needed to predict the job category is fairly low, with the exception of job category 2. The MAPE score is also higher for this category. The explanation might be that the jobs found in the second category are typically auxiliary jobs for the main business of the organization, thus employees in this category represent the most diverse group in terms of responsibilities and skills.

Cluster	k	MAE	MAPE	Nb. M+	Nb. M-	Δ Nb.
Job cat. 1	7	35.02	14.6%	34	35	-1
Job cat. 2	14	40.19	19.0%	41	31	+10
Job cat. 3	5	43.16	13.7%	82	88	-6
Job cat. 4	4	25.92	10.3%	118	131	-13
Job cat. 5	3	35.02	14.6%	30	36	-6
Job cat. 6	4	41.72	16.6%	35	38	-1

Table 4. k-NN Quality Metrics for Men's Training Set

Cluster	Nb. W+	Nb. W-	Δ Nb.	Mean $\Delta W+$	Mean $\Delta W-$
Job cat. 1	26	16	+10	21%	-14%
Job cat. 2	71	35	+36	23%	-14%
Job cat. 3	133	74	+59	28%	-12%
Job cat. 4	179	162	+17	14%	-11%
Job cat. 5	8	11	-3	13%	-10%
Job cat. 6	56	33	+23	28%	-16%

Table 5. k-NN Performance Statistics when Applied on Women

Table 5 presents the results of predicting women salary with the k-NN models trained on men. The table reports for each job category the number of women that were predicted by the algorithm to have a higher salary than the actual one (Nb. W+), smaller than the actual one (Nb. W-), the difference between the Nb. W+ and Nb. W- columns (Δ Nb.), the mean percentage of additional salary suggested by the algorithm (Mean Δ W+) , the mean percentage of the extraneous salary as identified by the algorithm (Mean Δ W-).

Except for the fifth category, the model show that Δ Nb., the difference between the number of women to whom a raise is suggested and the number of women to whom a salary decrease is suggested, is positive oftentimes greater than 10. By contrast, for the men, the model is fairly balanced with five cases out of six where Δ Nb is negative and much smaller in absolute value. Moreover, we observe that the mean percentage of additional salary is greater than the mean percentage of extraneous salary, which confirms that overall the algorithm suggests that women are more underpaid than they are overpaid. We see that job categories 2, 3 and 6 are those that are associated with the greatest level of inequality. Interestingly enough we also notice that in job categories 4 and 5, which group together positions that are closely related to the core business of the organization, we do not have such important differences.

5.3 Effect of Adjusting Salaries

In an ideal world, the company that identifies these pay gaps also has access to enough budget to correct them at once. In this section we assume that this is indeed the case. Following the results from the previous section, we are interested in understanding what would be the overall situation of the company if it were to implement the salaries adjustments from the k-NN predictions. First, we consider adjusting men and women according to the k-NN algorithm trained on the male population, regardless of whether the algorithm suggests a salary raise or salary cut. In the second scenario, we only modify the salaries for women though again adjusting up and down. In the last scenario we only raise women's salaries and leave all the other one unchanged. This latter scenario is naturally the most realistic one. For evaluation of the results, we compare the overall performance of the organization as we reported in Table 3.

The output from the first scenario is presented in Table 6. We observe that the Gini index which is considerably lower for both men and women as opposed to the initial results presented in Table 3. This is especially important for the men in the categories where the difference between the Gini index of the men's salaries and the Gini index of the women's salaries were greater at the start. While, based on the difference in mean and median salaries between men and women, the pay gap lowers (from 14% to 10% for mean salary and from 12% to 9.4% for median salary), we still observe that ASD holds in every cluster though the parameter ϵ decreases slightly for all but job category 2.1 an job category 1.2. In some sense, the overall situation appears mildly worse than it was regarding the salary distributions. Initially there was one category where the salary distribution of men did not dominate the salary distribution of women. Those results are not surprising as our hypothesis was that men's salaries were the reference to which we sought to align women's salaries. By adjusting both men and women salaries, we are not consistent with this hypothesis.

Cluster	Gini men	Gini women	Δ mean	Δ median	ASD	ϵ
Job cat. 1.1	0.15	0.12	17%	**23%**	**YES**	0.99
Job cat. 1.2	0.11	0.08	**4%**	2%	**YES**	0.96
Job cat. 2.1	0.07	0.08	4%	-3%	**YES**	0.95
Job cat. 3.1	0.2	0.14	10%	1%	**YES**	0.99
Job cat. 3.2	0.18	0.15	14%	13%	**YES**	0.99
Job cat. 3.3	0.16	0.14	17%	12%	**YES**	0.99
Job cat. 4.1	0.14	0.09	13%	**5%**	**YES**	0.99
Job cat. 4.2	0.12	0.09	4%	**2%**	**YES**	0.99
Job cat. 4.3	0.14	0.09	24%	**31%**	**YES**	0.99
Job cat. 5.1	0.13	0.11	7%	4%	**YES**	0.99
Job cat. 6.1	0.12	0.1	1%	2%	**YES**	0.54
Job cat. 6.2	0.14	0.13	0%	10%	**YES**	0.61
Average	0.14	0.11	10%	9.4%		0.92

Table 6. Corrected Cluster Gap Metrics under the First Scenario (Everybody + All Adjustments)

Therefore, in the second scenario, we only adjust women's salaries. The resulting organizational performance is reported in Table 7. The Gini index is lower for women in comparison with the baseline, but higher than the one presented in the first scenario which

was the results of an overfit of the model. In this scenario, as we expected, we see an improvement in terms of ASD, with the complete elimination of the dominance for one of the clusters and a minor reduction at the average ϵ level. While the delta mean grows for most of the clusters, the delta median lowers. The interpretation of this result is that while for lower paying women we achieve greater equality, there is a bigger gap for those in better paying positions.

Cluster	Gini men	Gini women	Δ mean	Δ median	ASD	ϵ
Job cat. 1.1	0.22	0.12	19%	6%	**YES**	0.98
Job cat. 1.2	0.14	0.08	4%	4%	**YES**	0.88
Job cat. 2.1	0.15	0.09	6%	3%	**YES**	0.87
Job cat. 3.1	0.22	0.14	14%	3%	**YES**	0.99
Job cat. 3.2	0.2	0.15	12%	0%	**YES**	0.96
Job cat. 3.3	0.17	0.14	19%	16%	**YES**	0.99
Job cat. 4.1	0.17	0.09	20%	8%	**YES**	0.99
Job cat. 4.2	0.15	0.09	7%	0.9%	**YES**	0.97
Job cat. 4.3	0.18	0.09	26%	19%	**YES**	0.99
Job cat. 5.1	0.17	0.11	13%	0.03%	**YES**	0.99
Job cat. 6.1	0.13	0.1	-0.2%	-0.2%	NO	
Job cat. 6.2	0.17	0.13	10%	17%	**YES**	0.90
Average	0.16	0.13	11.3%	8.6%		0.78

Table 7. Corrected Cluster Gap Metrics under the Second Scenario (Women + All Adjustments)

In practice we can only take into consideration the cases where the salary is increased from the baseline. The associated results are presented in Table 8. The corrections have a positive impact on bridging the gap between the salaries of the two categories. Indeed, fewer clusters (eight vs. twelve initially) display almost stochastic dominance between men and women and again we observe a decrease in the ϵ parameter for each cluster where ASD still holds. Note that even though the women's salaries are adjusted abased on the men's salaries, which displayed a higher Gini index, this effect does not translate to the women's group. Specifically, the Gini index in all the clusters is lower for the women by comparison to the initial results. Both the delta of mean and median salary have decreased, with the median indicating somehow an advantage of women over men for the first time yet the absolute magnitude of the advantage is much smaller than the starting point.

Interestingly enough, there is still a difference between the mean salary for women and for men. This might be due to factors that go beyond equal pay, such as the number of women in management position and that benefit from talent programs. Altogether, that third scenario offers some quite promising perspective, leading to an effective narrowing of the gender pay gap.

5.4 Catch-up Plan Under Budget Constraints

In the previous sections, we have presented our results under the hypothesis that there was sufficient budget to cover all the costs related to the necessary salary adjustments. In this section, we revisit this hypothesis and propose an optimization strategy that adapts to limited resources. Suppose that our costs are the differences between the actual salary and the predicted salary and for illustration purposes, our budget is able to cover only half of the total cost. The simple solution would be to give everybody half of the suggested increase. However, this approach might not be equitable as clusters vary in terms of level of pay inequality. We thus turn to more sophisticated optimization approaches.

Cluster	Gini men	Gini women	Δ mean	Δ median	ASD	ϵ
Job cat. 1.1	0.22	0.14	10%	1%	**YES**	0.96
Job cat. 1.2	0.14	0.14	-5%	-2%	NO	
Job cat. 2.1	0.15	0.1	-1%	-4%	NO	
Job cat. 3.1	0.20	0.15	8%	-1%	**YES**	0.97
Job cat. 3.2	0.23	0.15	7%	-5%	**YES**	0.91
Job cat. 3.3	0.19	0.14	11%	9%	**YES**	0.97
Job cat. 4.1	0.18	0.11	11%	0%	**YES**	0.99
Job cat. 4.2	0.16	0.12	0%	-3%	**YES**	0.66
Job cat. 4.3	0.19	0.1	21%	20%	**YES**	0.98
Job cat. 5.1	0.19	0.12	7%	-2%	**YES**	0.99
Job cat. 6.1	0.16	0.12	-9%	-6%	NO	
Job cat. 6.2	0.2	0.18	0%	7%	NO	
Average	0.18	0.13	5.3%	0.1%		0.69

Table 8. Corrected Cluster Gap Metrics under the Third Scenario (Women + Increases Only)

Specifically, we frame the problem as a knapsack problem, a standard problem in combinatorial optimization which can be nicely transposed to our settings. In the knapsack problem, we have on one side a sack (i.e., backpack) with limited capacity in terms of total weight and on the other side a set of items, each with a different value and weight. The knapsack algorithm ensures that we select the items that are the most valuable to us, while fulfilling the maximum weight constraints. In our situation, the value of each item is the improvement in pay equity (which we suggested to measure through the change in median salary difference between men and women), the weight corresponds to the salary increase cost. For our case we are going to consider a variation of this problem where we are able to divide the items which is called the fractional knapsack problem. This setting is more appropriate for us because we chose to allow for partial salary increases.

Cluster	Gains	Total Cost	Mean Cost	Fraction Total	Fraction Mean
Job cat. 1.1	20%	549	36	1	1
Job cat. 1.2	10%	276	25	1	1
Job cat. 2.1	11%	1920	32	0	0.03
Job cat. 3.1	22%	1841	37	1	1
Job cat. 3.2	20%	2472	46	0.4	1
Job cat. 3.3	5%	885	28	0	0
Job cat. 4.1	3%	1395	20	0	0
Job cat. 4.2	2%	1195	16	0	0
Job cat. 4.3	6%	475	14	1	1
Job cat. 5	6%	241	24	1	0
Job cat. 6.1	12%	1260	39	1	0
Job cat. 6.2	13%	949	39	1	1

Table 9. Optimal Distribution of Resources under Budget Constraints - Cluster Level

We undertook three sets of experiments, two where the items are clusters and one where the items are individuals. In the first cluster level experiment, the cost associated with each cluster is the total amount of money required for the increases while in the second cluster level experiment, the cost represents the average amount of money needed in each job category. For the experiment done at individual level, the cost is defined simply as the difference between the actual salary and the predicted salary, while the gains were set to be the same as the cluster level, for every women, according to the cluster they belong to. The constraints set for this experiment is half the budget needed completely close the equal pay gap.

In Table 9 we present the results for the two cluster level optimizations. The penultimate column reports for each job category the optimal level of salary increases based on total cost while the last column is for the experiment based on mean cost. We see that for both cases, the results yield only one cluster with fractional allocation of salary increases. In other words, the optimization algorithm is equivalent to ranking the clusters by order of return for investment and allocating a partial salary to the last cluster that can be included so as to fulfil the budget constraints. Intuitively, the algorithms promotes in the first scenario the clusters that have lower overall costs. However, this heuristic is faulty, because clusters with large number of women are left unchanged. Finally, Table 10 reports the results from the application of the knapsack algorithm at the level of an individual. Specifically, for each job category, we report the number of women who would receive the full increase (#1), partial increase (#fraction) or no increase at all (#0). This table highlight the suboptimality of reasoning at cluster level as for each of the job categories, we find women either receiving full increase or no increase, leading to much a finer level of adjustment. As a consequence, we would recommend using the third configuration which leads to what we feel is a more just allocation of scarce resources.

Cluster	#1	#fraction	#0
Job cat. 1.1	12	0	2
Job cat. 1.2	7	0	3
Job cat. 2.1	41	0	16
Job cat. 3.1	44	0	3
Job cat. 3.2	45	1	6
Job cat. 3.3	10	0	18
Job cat. 4.1	32	0	33
Job cat. 4.2	29	0	43
Job cat. 4.3	26	0	5
Job cat. 5	5	0	5
Job cat. 6.1	25	0	7
Job cat. 6.2	19	0	5

Table 10. Optimal Distribution of Resources under Budget Constraints - Individual Level

6 CONCLUSION

In this paper we present methodology leveraging artificial intelligence, statistics and operations research designed (i)to assess the status of an organization with respect to equal pay and also (ii)to recommend a viable plan whenever relevant. This method is currently being piloted by the HR department of an entity with more than 1500 people. we have purportedly chosen to use straightforward and easy to interpret machine learning models (as opposed to complex black box models), giving the HR department the liberty to make their own judgements and explain to non-technical managers the decisions taken. We hope that our approach empowers HR departments to develop an actionable yet disciplined approach to equal pay. Although we have applied this method to address differences in terms of compensation between men and women, it could be easily transferred to other types of discrimination, such as race, religion, national origin, or disability.

REFERENCES

[1] PC Álvarez-Esteban, Eustasio del Barrio, Juan Antonio Cuesta-Albertos, C Matrán, et al., 'Models for the assessment of treatment improvement: The ideal and the feasible', *Statistical Science*, **32**(3), 469–485, (2017).

[2] Marianne Bertrand and Kevin F Hallock, 'The gender gap in top corporate jobs', *ILR Review*, **55**(1), 3–21, (2001).

[3] Rotem Dror, Segev Shlomov, and Roi Reichart, 'Deep dominance-how to properly compare deep neural models', in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 2773–2785, (2019).

[4] Barry Gerhart, 'Gender differences in current and starting salaries: The role of performance, college major, and job title', *ILR Review*, **43**(4), 418–433, (1990).

[5] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 2nd edn., 1989.

[6] Ben Jann, 'The blinder–oaxaca decomposition for linear regression models', *The Stata Journal*, **8**(4), 453–479, (2008).

[7] Debra Leaker et al., 'The gender pay gap in the uk', *Economic & Labour Market Review*, **2**(4), 19–24, (2008).

[8] Robert I Lerman and Shlomo Yitzhaki, 'A note on the calculation and interpretation of the gini index', *Economics Letters*, **15**(3-4), 363–368, (1984).

[9] Claudia Olivetti and Barbara Petrongolo, 'Unequal pay or unequal employment? a cross-country analysis of gender gaps', *Journal of Labor Economics*, **26**(4), 621–654, (2008).

[10] UN Women. Heforshe 2019 impact report. https://www.heforshe.org/sites/default/files/2019-09/HeForShe%202019%20IMPACT%20 Report$_F ull.pdf$, 2019. $Accessed$: $2019 - 11 - 21$.

[11] Elaine L Zanutto, 'A comparison of propensity score and linear regression analysis of complex survey data', *Journal of data Science*, **4**(1), 67–91, (2006).