

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

DISCRIMINATION DETECTION AND OBSTRUCTION IN DIVERSITY WORKFORCE MONITORING IN EMPLOYMENT USING DATAMINING TECHNIQUES

Baskaran P¹, Dr. Arulanandam K².

Research scholar, Dept of Computer Science., Manonmanium Sundaranar University., Tirunelveli, Tamilnadu, India¹

-Corresponding author

Research Supervisor, HOD and Professor, Dept. of Computer Applications, G.T.M Govt. Arts and Science College,
Gudiyatam, Tamilnadu, India²

Abstract: This paper discusses the process of discrimination prevention by Diversity monitoring is the procedure through which a public or private entity collects stores and analysis data about the masterpiece of its workforce from corner to corner the related equality grounds. Monitoring is one of the most competent measures an employer can take to ensure it is in fulfillment with the equal opportunities. Several if not most employers eagerly collect and store data about the employees, in particular information relating to their sex, address, union membership, length of service and other data as may be required to pay the salaries and manage the workforce in general. Diversity monitoring basically implies extending the scope of information that is being collected to include information relating to the equality justification. Where monitoring practices exist, data is most commonly collected in relation to ethnic origin and disability, although there is at least some experience of monitoring across all of the equality. While it therefore appears to be theoretically and technically possible to conduct monitoring across all the grounds, challenges remain with respect to the general adequacy, and therefore this paper deals with a process for measuring diversity monitoring and discover discrimination and realize prevention method in workforce monitoring on employee.

Keywords: anti-discrimination, knowledge discovery, Work Force Monitoring, datasets, prevention

I. INTRODUCTION

Data is typically calm by means of questionnaires, interviews or all the way through surveillance. Generally speaking, data can be composed for two purposes that should not be confused. For the purposes of being able to draw conclusions and/or make decisions on a group level, for instance, data may reveal patterns of under-representation in particular occupations or educational institutions, and these may subsequently be available to back up an indirect discrimination claim or to plan a positive action strategy. The latter purpose highlights what statistics are about. The purpose of statistics is to bring to light, in an aggregate form, information about a certain population (e.g. the average employment rate of young immigrants) or a certain phenomenon (e.g. the number of discrimination complaints filed within a year). The purpose is to produce consolidated and anonymous information; but while the information is made – or at any rate: should be made available in such a manner that a particular person and a particular data cannot be related to each other, this does not preclude the need to observe data protection laws in the course of production of the statistics, as the aggregate data is always based on micro-data, typically personal data.

II. DATA COLLECTION

Data collection is relevant in respect of several key aspects of the directions:

- The concepts of discrimination, particularly indirect discrimination;

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

- The positive action provisions, which permit the maintenance or adoption of specific measures to prevent or compensate for disadvantages linked to any of the grounds mentioned in the Directives;
- The requirements to promote social dialogue between the two sides of industry with a view to fostering equal treatment, including through the monitoring of workplace practices.

III. DISCRIMINATION IN EMPLOYMENT

“Discrimination in employment and occupation” refers to practices that have the effect of placing certain individuals in a position of subordination or disadvantage in the labor market or the workplace because of their race, color, religion, sex, political opinion, national extraction, social origin or any other attribute which bears no relation to the job to be performed. Discriminatory practices can be direct or indirect. Direct discrimination arises when an explicit distinction, preference or exclusion is made on one or more grounds. For example, a job advertisement for “men only” would constitute direct discrimination.

Indirect discrimination refers to situations, measures or practices that are apparently neutral but which in fact have a negative impact on persons from a certain group. The latter type of discrimination, because of its more hidden nature, is the most difficult to tackle.

Equality of opportunity and treatment allows all individuals to fully develop their talents and skills according to their aspirations and preferences, and to enjoy equal access to employment as well as equal working conditions. To achieve full freedom from discrimination in employment and occupation, the mere removal of discriminatory practices does not suffice. It is also necessary to promote equality of opportunity and treatment in the workplace at all stages of the employment relationship, including recruitment, retention, promotion and termination practices, remuneration, access to vocational training and skills development.

IV. NON-DISCRIMINATORY SELECTION DECISION

For example, A Candidate with black skin color called ‘X’ applies for an executive position with the employer, a health maintenance organization. X is well qualified; he has a B.S. in biology from a large state university and an M.D. from a prestigious private university. X also has seven years experience practicing internal medicine and recently obtained an Executive M.B.A. from a well-respected business school. The employer interviewed X and eight other candidates. And X was one of two finalists brought back for a final round of interviews. The employer’s selection committee ultimately chose other Candidate CALLED ‘Y’ with white skin color, a finalist with slightly fewer qualifications but with experience in a similar job for a competitor. The employer tells to HR that given the selected candidate Y’s experience, it believed it would gain the most competitive benefit by hiring him. The HR investigator confirms Y’s experience working for a competitor, and reads the minutes of the selection committee’s final meeting which reflect that this was the reason discussed at the meeting for choosing Y over X. Here, the evidence supports the employer’s legitimate, nondiscriminatory reason.

V. DISCRIMINATORY SELECTION DECISION

X, files a charge after he applied for a promotion, was interviewed, and was not selected. The investigation reveals that, based on objective qualifications, X was deemed one of the top candidates but the job ended up going to Y, a similarly qualified candidate from outside the company. The hiring manager tells the investigator that he thought that X was well qualified but he chose Y because he “seemed to be a better fit; I’m comfortable with him and I can see him in my job one day.” When pressed to be more specific, the manager says he liked the fact that Y worked for a competitor. However, the investigation reveals that although Y did work for another company in the industry, it was not really a competitor. Employee and management witnesses tell the investigator that Y’s experience working for another company in the industry was no more valuable than X’s experience working for the company itself. The witnesses also tell the investigator that, until now, the company practice had been to prefer qualified internal candidates over similarly qualified

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

external candidates. There is reasonable cause to believe that X was discriminated against based on his some of the reasons which already discussed in types of discrimination

1. Failure to Promote

X, an administrative assistant, filed a charge alleging that she receives less pay than the office manager even though in her opinion they perform similar work. The investigator concludes that X is not similarly situated to the office manager due to the difference in responsibility associated with the jobs. Nevertheless, the investigation reveals that all but one of R's employees hold lower paying clerical, secretarial, and low-level administrative positions. Many of these employees testified to the lack of promotional opportunities into higher paying jobs. R asserted that it does not employ in higher paying jobs because of a lack of qualified applicants. The investigator determines that qualified employees have applied for these jobs but nearly all, like X, have not been promoted.

2. Different terms and conditions

X, works for a computer software company. The company thrives on active socializing between employees and decision makers both on and off the job – from lunch outings, after-work happy hours and weekend golf outings, to children's birthday parties and family barbecues. Many employees establish strong relationships with decision makers through these informal networks, and as a result, tend to get put on the plum projects and get the plum promotions. X has experienced difficulty in building relationships with decision makers because he often receives invitations late or indirectly from peers, rather from the decision makers themselves. After being passed over for several important projects, X files charge alleging race/national origin discrimination because he believes he is being excluded from his workplace network for reasons related to his other descent. X's exclusion would be actionable if it affects the terms and conditions of his employment

3. Retaliation

X has felt like he has been treated less favorably than his female co-workers. X believes his supervisor acts discriminately towards men and files a complaint with the Human Resources department and then another complaint with Metro. Shortly after his employer receives his Metro complaint, X is terminated for "not being a team player." X should then return to Metro to file a complaint of retaliation. X has engaged in the statutorily protected activity of protesting discriminatory treatment. Using these same circumstances, one of X's co-workers provides the Metro investigator with evidence showing their supervisor's bias towards men. Shortly after this co-worker's interview with the investigator, the supervisor cuts this employee's hours. The co-worker is also protected under the laws

4. Compensation

A male tennis instructor (x) and a female tennis instructor (y) at a particular health club provide tennis lessons that are substantially equal. The x instructor is paid a weekly salary, but the y instructor is paid by the lesson. Even if the two instructors receive essentially the same pay per week, there is a violation because the male and female are not paid in the same form for substantially equal work.

5. Training, advancement, opportunities

X, is a new office clerk. Her primary duties are to sort and file purchase orders and invoices. Within a few weeks, it is clear to the employer that X is processing her purchase orders and invoices too slowly due to mistakes. The employer

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

terminates X, who then files a charge alleging race discrimination. The investigation reveals that although White employees who perform at a substandard level are coached toward increasingly good performance, X and other employees of color get less feedback and thus tend to repeat mistakes and make new ones that could have been avoided. The evidence establishes that the employer unlawfully terminated X.

6. Discipline and Discharge

X, a retail store clerk, is frequently 10-15 minutes late for her shift on several days per week when she attends some family occasion across town. Her manager, Y, has never disciplined her for this tardiness, and instead filled in for her at the cash register until she arrived, stating that he understood her situation. On the other hand, Z, a newly hired clerk, is disciplined by Y for arriving 10 minutes late for his shift even though Y knows it is due to his some domestic works. While Y can require all similarly situated employees to be punctual, he is engaging in disparate treatment based on some reason by disciplining only Z and not X absent a legitimate nondiscriminatory reason for treating them differently.

7. Employment Decision Based on Customer Preference

X, is hired to work at the counter in a coffee shop. A few weeks after X begins working, the manager notices that the work crew from the construction site near the shop no longer comes in for coffee in the mornings. When he inquires, the crew complains that X, whom they mistakenly believe is belongs to some notorious people, makes them uncomfortable in that place. The manager tells X that he has to let him go because the customers' discomfort is understandable. The manager has subjected X to unlawful religious discrimination by taking an adverse action based on customers' preference not to have a cashier X's perceived persons. X's termination based on customer preference would violate regardless of whether he was any other peoples like strangers or notorious peoples.

VI. DIVERSITY MONITORING AND MEASURING DISCRIMINATION ON EMPLOYMENT

Monitoring recruitment and selection Employers interested in ensuring that their recruitment and selection practices are not biased need to address the following questions:

- Do qualified members of all groups apply for advertised posts in proportion to their presence in the population?
- Given the population of qualified applicants who do apply, does each group have the same chance of getting on the shortlist?
- Given the applicants on the shortlist, does each group have the same chances of getting offered the job?

Data external to the organization is required to answer the first question. Proportions of applicants from different protected groups need to be compared with their proportions of the qualified population in general. For that purpose, employers need to know the likely pools of suitably qualified applicants in the relevant spatial labor market. Where the distributions of applicants, offers or hiring outcomes by particular groups deviate, in a statistically significant sense, from their distribution in the relevant qualified populations, then discrimination may be present and this possibility needs further investigation.

Even where it is not possible in a satisfactory manner to investigate whether the pool of applicants is representative of the qualified general population, the monitoring of the applicants can trace irregularities in the process by which applicants are short-listed, invited to interview and offered a job. Diversity monitoring does thus not become redundant in the face of possible lack of suitable external comparative data. It is crucial that the employer ties monitoring to concrete remedial action. The type of corrective action should directly address the source of the problem as identified by the analysis of monitoring data.

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

1. Workforce monitoring

Once individuals are working within an organization, it is not difficult to keep records of all employees. If the relevant data has already been collected at the recruitment stage, employer needs only to transfer the data concerning appointed individuals into the employee database, while the data on other applicants may for instance be rendered anonymous and kept for a fixed period of time in order for the employer to be able to respond effectively to any claims of discrimination in relation to the recruitment process.

2. Measuring Discrimination – Data Collection

Workforce diversity monitoring should ideally cover all aspects relating to employment, including promotion, pay and other conditions of work, and termination of employment relationships. It is of high importance to monitor how individuals progress to the top levels of jobs. Where employers are monitoring progression, the appropriate match would be, for instance, the success rates of each ethnic group. It may also be necessary to collect data in relation to disability and the need for reasonable accommodation. No external comparative data is necessarily needed to assess equality of treatment at this stage.

3. Technical and practical considerations

In principle, the within-organization elements of diversity monitoring are not inherently problematic or technically challenging. Basically, applicants for vacancies are for instance asked to fill in an equal opportunity form, usually as a voluntary exercise. To alleviate any fears of misuse of the data in the selection process, the processing of this information can be separated from the processing of the applications proper and trusted to a person who is not involved in the selection process. Once information about short-listing, appointments, salary and promotion become available, these can be entered into the database. The overall statistics on applicants and the workforce, broken down by different equality grounds, can then be compared with the relevant benchmark data.

In practice, however, things are not this simple. Many employers feel uncomfortable about asking questions about sensitive issues, and many prospective employees feel uncomfortable about answering those questions. Practice has shown that where collected on a voluntary basis, the information on ethnicity, disability and sexual orientation is often incomplete. Missing data has a direct and detrimental effect on the quality of the monitoring exercise, which is why employers need to consider ways in which they can encourage employees to submit the data. A major technical challenge relates to the existence and availability of suitable external data, which can be beneficial at some stages of monitoring for comparative purposes. An employer or a service provider needs a benchmark, a standard, in order to be able to assess its internal data. Where such a benchmark based on official data exists, it provides clear benefits for the monitoring exercise. An ideal benchmark would show the proportion of qualified applicants belonging to a certain protected category living in the spatial area of recruitment.

This requires the existence of extensive and reliable macro-level data sets that include personal data in relation to the protected equality ground(s), qualifications (in particular highest education) and place of residence. Such data could even in theory be only available through censuses, central population registers or possibly national Labor Force Surveys. In situations where official data is not available, data from similar organizations may provide for a useful point of comparison. Comparative analysis can reliably be carried out only where the categories used for collecting data through monitoring correspond to those used in the official datasets (or in the other benchmark used). For instance, it is clear that if the monitoring form used by an employer defines 'disability' only in terms of physical impairment, while the relevant official data is based on a more inclusive definition, the two do not provide for a common basis of comparison. As regards the

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

spatial area of recruitment, it has to be noted that employers use different recruitment strategies for different types of vacancies. Local outlets are typically used for low-grade or starter entry jobs and national outlets are used for higher-grade entry jobs. Some of the jobs advertised are so specialized that it is practically impossible to ever hope to gain information from nationally representative surveys or even census data about the pool of qualified applicants.

VII. ALTERNATIVE METHOD: ANONYMOUS MONITORING

Under-declaration, especially of ethnic, religious and sexual identity causes many difficulties for the analysis of the data. One way in which this problem can be addressed is the use of anonymous monitoring instead of non-anonymous monitoring. Collection of data through anonymous surveys, whereby no data on identified or identifiable persons is gathered, has been found to significantly increase response rates especially among the LGB people. If repeated at regular intervals, and provided that the survey sample is representative, anonymous surveys can yield also trend data.

Whether anonymous monitoring has any benefits over non-anonymous monitoring in terms of response rates and anonymity depends on the size of the organization and the level of information that is being sought. In a small or even medium-size organization, a detailed questionnaire may lead to a situation where particular responses can never be traced back to particular individuals, a fact which compromises the very idea of anonymous monitoring and may again lead to a decrease in the willingness to co-operate. The usefulness of this method has thus to be assessed on a case-by-case basis.

The disadvantage of relying on regular anonymous surveys instead of employee registers is that the information cannot be kept up-to-date, in addition to which it is a bit more burdensome method both for the employees and the employer. Unlike anonymous monitoring, non-anonymous monitoring may be useful for the exploration of reasonable accommodation duties and for the implementation of equality (positive action) plans. Anonymous monitoring also faces the same challenges associated with the need for suitable comparative statistics as non-anonymous monitoring. Overall, it may be concluded that non-anonymous monitoring has some advantages over anonymous monitoring, and that the benefits of anonymity may not materialize where detailed information is needed or where the size of the company or other entity is small. Anonymous monitoring may however be the only available option where national data protection laws do not allow the use of non-anonymous monitoring or where there are serious problems with response rates.

VIII. DIVERSITY MONITORING IN SERVICE DELIVERY

The prohibition of discrimination under the Racial Equality Directive applies not just in relation to employment, but also in relation to, education, social security, healthcare, social advantages and access to and supply of goods and services, including housing. Entities working in these areas may thus want to monitor not just their personnel but also their service delivery. Monitoring can provide useful information particularly in the domains of housing, education and health care. Monitoring can be used to tell which groups are using or receiving the services and whether certain groups are facing particular challenges. An agency providing housing services may for instance want to draw up profiles of service use to verify that all groups have an equal access to housing of equal quality and on equal terms. Schools can assess the impact of their policies on pupils, staff, and parents from different ethnic groups. They can also monitor the way their policies and operating modes work, with special emphasis on pupils' attainment levels. Further and higher education institutions can monitor, by ethnic group, student admissions and progress, and staff recruitment and career development. This is understandable, given that most LGB people are thought not to be out about their sexual orientation to everyone they work with. It should be noted that different data protection rules may apply where the data is used not just for monitoring purposes but also for the purposes of making decisions with respect to specific individuals. Another way to deal with low response rates may be to use other-identification in addition to self-identification as the basis on which assignment to different categories takes place.

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

IX.DISCRIMINATION DISCOVERY

A dataset is a collection of data items (report/records) and their attributes.

- Let D be the original dataset. An item is an attribute along with its value, e.g. Race=black.
- An itemset, i.e. A , is a collection of one or more items, e.g. {Foreign worker=Yes, City=NYC}.
- A classification rule is an expression $A \rightarrow C$, where C is a class item (a yes/no decision), and A is an itemset containing no class item, e.g. {Foreign worker=Yes, City=NYC} Hire=no.
- A is called the premise of the rule. A frequent classification rule is a classification rule with a support or confidence greater than a specified lower bound.

"Let DTs be the set of predetermined discriminatory items in D (e.g. $DTs = \{\text{Foreign worker=Yes, Race=Black, Gender=Female}\}$). Frequent classification rules fall into one of the following two classes:

- 1) A classification rule $A \rightarrow C$ is potentially discriminatory (PD) when $A = X, Y$ with X DTs a non-empty discriminatory itemset and Y a non-discriminatory itemset. For example {Foreign worker=Yes; City=NYC} Hire=No.
- 2) A classification rule $A \rightarrow C$ is potentially non-discriminatory (PND) when $A = D, Y$ is a non-discriminatory itemset. For example {Zip=10451, City=NYC} Hire=No, or {Experience=Low; City=NYC} Hire=No.
- 3) As mentioned before, the translated qualitative statements in existing laws, regulations and legal cases into quantitative formal counterparts over classification rules and they introduced a family of measures of the degree of discrimination of a PD rule for direct discriminatory discovery and a PND.
- 4) For indirect discrimination discovery. Then, whether the PD rule is potentially directly discriminatory can be assessed by threshold. Based on this measure, PD rules could be discriminatory or protective. In addition, whether the PND rule is potentially indirectly discriminatory can be assessed by threshold. Based on this measure, PND rules could be redlining or non-redlining (legitimate).

X.DISCRIMINATION PREVENTION

Beyond discrimination discovery, preventing knowledge-based decision support systems from making discriminatory decisions (discrimination prevention) is a more challenging issue. It can be even more difficult when we want to prevent not only direct discrimination but also indirect discrimination or both at the same time. The classification of discrimination prevention methods is related to the way of eliminating discrimination and also to the phase of the data mining process in which discrimination prevention is done. Based on this criterion the discrimination prevention methods fall into three groups:

• Pre-processing:

Transform the source data in such a way that the discriminatory. Biases contained in the original data are removed so that no unfair decision rule can be mined from the transformed data and apply any of the standard data mining algorithms. The pre-processing approaches of data transformation and hierarchy based generalization can be adapted from the privacy preservation literature.

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

• In-processing:

Change the data mining algorithms in such a way that the resulting models do not contain unfair decision rules. For example, an alternative approach to cleaning the discrimination from the original dataset whereby the nondiscriminatory constraint is embedded into a decision tree learner by changing its splitting criterion and pruning strategy through a novel leaf relabeling approach. However, it is obvious that in-processing discrimination prevention methods must rely on new special-purpose data mining algorithms; standard data mining algorithms cannot be used.

• Post-processing:

Modify the resulting data mining models, instead of cleaning the original dataset or changing the data mining algorithms. For example, a confidence-altering approach is proposed for classification rules inferred by the CPAR algorithm. Although some methods have already been proposed for each of the above mentioned approaches (pre-processing, in-processing, post-processing), discrimination prevention stays a largely unexplored research avenue.

XI.CONCLUSIONS

The report concludes that there is considerable and compelling demand for collection of data in relation to the equality grounds to avoid discrimination. Accordingly, the report recognizes the need to engage in data collection in order to work towards the realization of equal treatment. As identified in this report, there is a wide range of contexts in which data collection is relevant and often necessary to prove and remedy discrimination. Equality bodies, must have the competence to carry out surveys on the extent and nature of discrimination. Where it is suspected that discrimination exists, for instance in that a particular group is not treated equally with respect to access to employment, education, housing or other services, states concerned are under an obligation to effectively investigate the issue, which may in practice require data collection. Socio-economic data, broken down by the equality grounds, is needed to document the disadvantages that the Member States are, under certain conditions, required to remedy by means of special measures. Individual complainants are in need of statistical data in order to successfully pursue their cases before courts and other competent bodies, especially in cases of indirect discrimination. It should however be noted that this does not necessarily mean that states would necessarily have to engage in collection of personal data; data collected by means of alternative methods, such as anonymous surveys, may meet many of the data needs identified above. This paper deals with a classification algorithms based rule is formed and implemented to a data collection for detecting discrimination and the same rule (PD rule) used for preventing the data sets from discrimination.

REFERENCES

- [1] UnitedStatesCongress, *US Equal Pay Act*, 1963. <http://archive.eeoc.gov/epa/anniversary/epa-40.html>
- [2] Parliament of the United Kingdom, *Discrimination Act*, 1975. http://www.opsi.gov.uk/acts/acts1975/PDF/ukpga_19750065_en.pdf
- [3]ParliamentoftheUnitedKingdom,*RaceRelationsAct1976*. <http://www.statutelaw.gov.uk/content.aspx?activeTextDocId=2059995>
- [4] European Commission, *EU Directive 2000/43/EC on Antidiscrimination*, 2000. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:180:0022:0026:EN:PDF>
- [5] D. Pedreschi, S. Ruggieri and F. Turini, "Discrimination-aware data mining". Proc. of the 14th ACM International Conference on Knowledge Discovery and Data Mining (KDD 2008), pp. 560- 568. ACM, 2012.
- [6] J. Natwichai, M. E. Orlowska and X. Sun, "Hiding sensitive associative classification rule by data reduction". Advanced Data Mining and Applications (ADMA 2007), LNCS 4632, pp: 310-322. 2007.
- [7] S. R. M. Oliveira and O. R. Zaiane. "A unified framework for protecting sensitive association rules in business collaboration". International Journal of Business Intelligence and Data Mining, 1(3):247287, 2006.
- [8] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases". Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487- 499. VLDB, 1994.
- [9]Agresti, A. Categorical data analysis (2nd ed.). Wiley-Interscience.. (2002).