# Analysis of Direct and Indirect Discrimination Discovery and Prevention Algorithms in Data Mining

Mr. Krishna Kumar Tripathi
Assistant Professor, Shivajirao Jondhale College of Engineering, Dombivali, Thane, Mumbai University.
Mumbai, Maharashtra, India

***Abstract*— In present world we are having collection of large amount of data. Data is increasing day by day. We need to analyze data to take some decision. Data Mining is a technique to retrieve fruitful results from large amount of data. When we apply data mining algorithms on data sets we are getting desired results. The results are based on rules and datasets there is no human interaction. But if dataset is biased then definitely discrimination decision can be taken. Discrimination means act of unfairly or unequal treating people on the basis of their gender, race, ideology etc. Discrimination divided into two category Direct and Indirect discrimination. Direct discrimination takes places when decisions are made based on sensitive attributes. Indirect discrimination takes place on non-sensitive attributes but these attributes strongly related with biased sensitive attributes. In this paper we focus on how to clean training data sets and outsourced datasets, such that no discrimination should occur. Also we discuss how to discover and prevent Direct and Indirect discrimination prevention techniques in Data Mining.**

***Keywords—Data Mining, Direct and Indirect discrimination, Rule protection, Rule generalization, Antidiscrimination.***

## I. Introduction

The word discrimination originates from the Latin discriminare, which means to “distinguish between”. In social sense, however, discrimination refers specifically to an action based on prejudice resulting in unfair treatment of people on the basis of their membership to a category, without regard to individual merit. As an example, U.S. federal laws [11] prohibit discrimination on the basis of race, color, religion, nationality, sex, marital status, age and pregnancy in a number of settings, including: credit/insurance scoring (Equal Credit Opportunity Act); sale, rental, and financing of housing (Fair Housing Act); personnel selection and wage (Intentional Employment Act, Equal Pay Act, Pregnancy Discrimination Act).[12]

On the other side, discrimination is rarely defined in rigorous and universal terms. First, protected-by-law groups, such as minorities and disadvantaged people, are sometimes not fully identified, leaving space for ambiguous issues such as in the debate about multiple, intersectional and compound discrimination discussed in ENAR (2007). Second, the interpretation of existing legislations leads to different quantitative measures of discrimination and, a fortiori, to different thresholds between what is legal and illegal. Third, discrimination can be hidden behind apparently neutral practices, known as indirect discrimination that must be unveiled by some deductive reasoning exploiting additional knowledge, which we call background knowledge. Fourth, a few policies, known as affirmative action’s, that favor minorities are allowed, encouraged or even enforced by laws. Finally, in case a prima-facie evidence of discrimination is found in the data, the anti-discrimination analyst has still to consider possible argumentations of the respondent, e.g., in opposing a genuine occupational requirement justification. We call these issues the deductive problem in discrimination discovery.[15] The first approach to discrimination discovery from a computer science perspective is based on extracting and reasoning about classification rules. The various concepts and analyses, originally implemented as a stand-alone program for achieving the best performances, have been re-designed around an Oracle database, storing extracted item sets and rules, and a collection of functions, procedures and snippets of SQL queries that implement the various legal reasoning for discrimination analysis[17][18][19]. Discrimination discovery is an interactive and iterative process, where analyses assume the form of deductive reasoning over extracted rules. An appropriately designed database, with optimized indexes, functions and query snippets, can be welcome by a large audience of users, including owners of Socially-sensitive decision data, government antidiscrimination analysts, technical consultants in legal cases, researchers in social sciences, economics and law. [16]

Discrimination can be either direct or indirect (also called systematic).Direct discrimination consists of set of laws(rules) or procedures(events) that explicitly mention minority or deprived groups based on sensitive discriminatory attributes related to group membership. Indirect discrimination consists of set of laws (rules) or procedures that, while not clearly mentioning discriminatory attributes, deliberately or not deliberately could generate discriminatory decisions. Redlining by financial institutions (refusing to grant mortgages or insurances in urban areas they consider as deteriorating) is an archetypal example of indirect discrimination, although definitely not the only one. With a slight neglect of language for the sake of compression, in this paper indirect discrimination will also be referred to as redlining and rules causing indirect discrimination will be called redlining rules [7]. Indirect discrimination could happen because of the availability of some background knowledge (rules) [1], because of the existence of nondiscriminatory attributes that are highly correlated with the sensitive ones in the original data set. The main charity of this paper is to provide the best solution for removing direct and/or indirect discrimination biases in the original data set while preserving data quality [20][21][22].

In this paper, Section II discusses related work; Section III introduces background and motivation; Section IV describe the scope of research in the direct and indirect discrimination prevention in data mining field ; Section V conclusions are made based on literature survey.

## II. Related Work

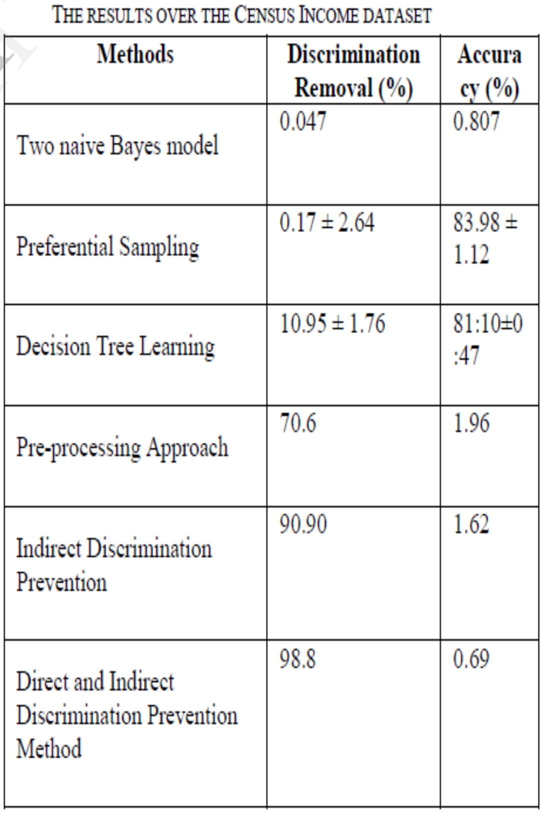
The discovery of discriminatory decisions was first proposed by Pedreschi et al. [12], [15]. The approach is based on mining classification rules (the inductive part) and reasoning on them (the deductive part) on the basis of quantitative measures of discrimination that formalize legal definitions of discrimination. Current discrimination discovery methods consider each rule individually for measuring discrimination without considering other rules or the relation between them [3]. Here we consider existence and non-existence discriminatory attribute. After discrimination discovery they focus on Discrimination prevention methods like :

- Preprocessing – First we transform the data so that discriminatory biases are removed. Then the preprocessing approaches of data transformation and hierarchy-based generalization can be adapted from the privacy preservation literature. The preprocessing approach is useful for applications in which a data set should be published and/or in which data mining needs to be performed also by external parties (and not just by the data holder).
- 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 data set is proposed in [2] 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 inprocessing discrimination prevention methods must rely on new special-purpose data mining algorithms; standard data mining algorithms cannot be used.
- Postprocessing. Modify the resulting data mining models, instead of cleaning the original data set or changing the data mining algorithms. For example, in [13], a confidence-altering approach is proposed for classification rules inferred by the CPAR algorithm. The postprocessing approach does not allow the data set to be published: only the modified data mining models can be published (knowledge publishing), hence data mining can be performed by the data holder only [3].

Also Pedreschi explained a proposal for direct and indirect discrimination prevention. In this section, we present our approach, including the data transformation methods that can be used for direct and/or indirect discrimination prevention. Our approach for direct and indirect discrimination prevention can be described in terms of two phases:

- Discrimination measurement. Direct and indirect discrimination discovery includes identifying discriminatory rules and redlining rules. To this end, first, based on predetermined discriminatory items in DB, frequent classification rules in FR are divided in two groups: PD and PND rules.
- Data transformation. Transform the original data DB in such a way to remove direct and/or indirect discriminatory biases, with minimum impact on the data and on legitimate decision rules, so that no unfair decision rule can be mined from the transformed data. In the following sections, we present the data transformation methods that can be used for this purpose. [3]

S.Subbulakshmi gives a comparison of state-of-the-art methods on the Census Income dataset. It affords the entitlement of discrimination and accuracy for the above discussed methods. The performances of various methods have been specified below Table 1 based on discrimination and accuracy [8][14].



THE RESULTS OVER THE CENSUS INCOME DATASET

| Methods | Discrimination Removal (%) | Accuracy (%) |
|---|---|---|
| Two naive Bayes model | 0.047 | 0.807 |
| Preferential Sampling | 0.17 ± 2.64 | 83.98 ± 1.12 |
| Decision Tree Learning | 10.95 ± 1.76 | 81:10±0 :47 |
| Pre-processing Approach | 70.6 | 1.96 |
| Indirect Discrimination Prevention | 90.90 | 1.62 |
| Direct and Indirect Discrimination Prevention Method | 98.8 | 0.69 |

Table 1: Comparison between different discrimination removal models[8]

In Toon Calders and Sicco Verwer paper they contribute as follows:

The discrimination-aware classification problem is illustrated and motivated.We show that simply removing the sensitive attribute from the training dataset does not solve the problem, due to the so-called red-lining effect. We propose three approaches to tackle the problem of discrimination-aware classification with Naive Bayes classifiers:

- in a post-processing phase we modify the probability of the decision being positive by changing the probabilities in the model,
- we train one model for every sensitive attribute value and balance them, and
- we add a latent variable in the Bayesian model that represents an unbiased, discrimination-free label and optimize the model parameters for likelihood using expectation maximization. We present and discuss experiments for the three approaches on both artificial and real-life data.[ 2]

Sara Hajian explained in their Discrimination Prevention in Data Mining for Intrusion and Crime Detection paper that The existing literature on anti-discrimination in computer science mainly elaborates on data mining models and related techniques. Some proposals are oriented to the discovery and measure of discrimination. Others deal with the prevention of discrimination.

- *Discrimination discovery* is based on formalizing legal definitions of discrimination1 and proposing quantitative measures for it. These measures were proposed by Pedreschi in 2008. This approach has been extended to encompass statistical significance of the extracted patterns of discrimination in, and it has been implemented as reported in. Data mining is a powerful aid for discrimination analysis, capable of discovering the patterns of discrimination that emerge from the data.
- *Discrimination prevention* consists of inducing patterns that do not lead to discriminatory decisions even if trained from a dataset containing them. Three approaches are conceivable: (i) adapting the preprocessing approaches of data transformation and hierarchy-based generalization from the privacy preservation literature [6], [10]; (ii) changing the data mining algorithms (in-processing) by integrating discrimination measure evaluations within them [11]; and (iii) post-processing the data mining model to reduce the possibility of discriminatory decisions [8]. Although some methods have been proposed, discrimination prevention stays a largely unexplored research avenue [5].

Faisal Kamiran explains Classification with No Discrimination . The CND method was implemented and tested on a credit score dataset displaying discriminatory behavior. Using our proposed CND method we were able to learn classifiers that no longer discriminate future data, without loosing too much accuracy. In summary, the contributions of this paper are as follows:

1) a formal definition of the non-discriminatory classification problem. This definition involves a measure for assessing the discrimination in a dataset,
2) a proposed solution, CND, for this problem, and
3) a performance study on a credit score dataset showing promising results.[7]

Also Faisal Kamiran gives two awareness techniques for the decision tree construction process:

- Dependency-Aware Tree Construction. When evaluating the splitting criterion for a tree node, not only its contribution to the accuracy, but also the level of discrimination caused by this split is evaluated.
- Leaf Relabeling. Normally, in a decision tree, the label of a leaf is determined by the majority class of the tuples that belong to this node in the training set. In leaf relabeling we change the label of selected leaves in such a way that discrimination is lowered with a minimal loss in accuracy.[9]

Salvatore Ruggieri introduce Dcube Architecture for Discrimination Discovery in Databases
DCUBE supports the discrimination discovery process shown in figure Fig. 1. The user starts the DCUBE wizard through the Oracle SQL Developer GUI. The wizard allows for selecting the following inputs:
(1) a relational table, view or SQL query from a JDBC data source, or from a CSV text file;
(2) a minimum support threshold;
(3) a list of PD items with all other items treated as PND;
(4) a list of class items;
(5) a target Oracle schema.
Additional inputs constraint the set of classification rules to be extracted by setting: the maximal size of a frequent itemset; the maximum support threshold; the maximal similarity threshold between items, after which two or more similar items are merged. Based on those inputs, DCUBE proceeds with the phases of mining, loading and querying [4], [16].

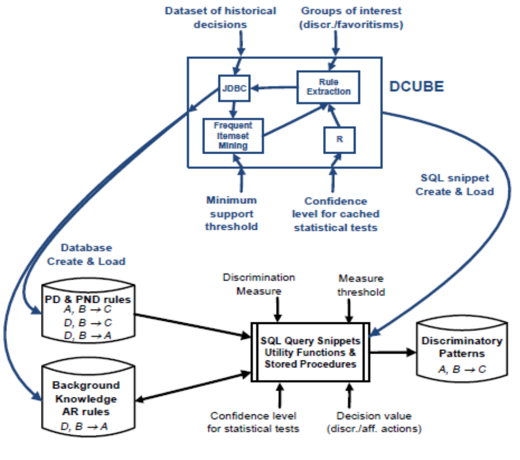

Dataset of historical decisions
Groups of interest (discr./favoritisms)
DCUBE
JDBC
Rule Extraction
Frequent Itemset Mining
R
Minimum support threshold
Confidence level for cached statistical tests
SQL snippet Create & Load
Database Create & Load
Discrimination Measure
Measure threshold
PD & PND rules
A, B → C
D, B → C
D, B → A
SQL Query Snippets Utility Functions & Stored Procedures
Discriminatory Patterns
A, B → C
Background Knowledge AR rules
D, B → A
Confidence level for statistical tests
Decision value (discr./aff. actions)

Figure [1]. Analysis process supported by DCUBE [16]

## III. BACKGROUNG AND MOTIVATION

Classification models are trained on the historical data for the prediction of the class labels of unknown data samples. Often, however, the historical data is biased towards certain groups or classes of objects. For example, throughout the years, in a certain organization black people might systematically have been denied from jobs. As such, the historical employment information of this company concerning job applications will be biased towards giving jobs to white people while denying jobs from black people. In order to reduce this type of racial discrimination, new laws requiring equal job opportunity have been enacted by the government. As such, the organization receives instructions in the form of, e.g., minimum quota for black employees. Suppose now that the company wants to partially automate its recruitment strategy by learning a classifier that predicts the most likely candidates for a job.[7]

The main motivation for starting out this research topic stems from a recently started collaboration with WODC; a Dutch study center associated with the department of Justice. The goal of this agency is providing data and modeling demographic and crime data to support policy making. Their interest emerges from the possibility of correlations between ethnicity and criminality that can only be partially explained by other attributes due to data incompleteness (e.g., latent factors). Learning models and classifiers on such data could lead to discriminatory recommendations to the decision makers. Removing the ethnicity attributes would not solve the problem due to the red-lining effect, but in contrast even aggravate it, as the discrimination still would be present, only it would be hidden better. In such situations our discrimination-aware data mining paradigm clearly applies; even though racial discrimination would improve the accuracy of our classifier, in the policy making context it is unacceptable.[2]

Like WODC we can take help of any research organization for Direct and Indirect discrimination prevention in data mining. Prevention can be done using several methods but still no one method is 100% effective. So there is scope of research in this Data mining area.

## IV. SCOPE OF RESEARCH

In this discrimination discovery, measure of discrimination and prevention of discrimination are the area of research. In current discrimination discovery methods we consider each rule individually for measuring discrimination without considering other rules or relation between them so in my research we are going to consider relation between rules for discrimination discovery based on existing or non-existence of discriminatory attributes.

Other scope of research is consider both cases direct and indirect discrimination instead of only direct discrimination, find their good tradeoff between discrimination removal and the quality of the resulting training data sets and data mining models. While removing discrimination, quality of data should be maintained.

There is large area of research in discrimination prevention field. There are some methods to prevent discrimination like Pre-processing, In-processing, Post-processing. In this research we first concentrate on discrimination prevention based on preprocessing, because the preprocessing approach seems the most flexible one: it does not require changing the standard data mining algorithms. Then we will focus on In-processing and post-processing methods to achieve effective results.

## V. CONCLUSION

In this paper, we explore how to discover Direct and Indirect discrimination also we have focus on prevention algorithms. Here we consider relation between discrimination discovery method rules for better results means we take into consideration relation between rules of discrimination discovery, based on the existence or nonexistence of discriminatory attributes. We perform data preprocessing to avoid direct and indirect discrimination. We first measure direct and indirect discrimination then we apply data transformation to remove discriminatory biases without violating data quality up to some extent. So there is research scope in direct and indirect discrimination preventions methods and we need to invent new methods for Data transformation.

## ACKNOWLEDGMENT

My sincere gratitude towards my professors of B.E (under graduate) and M.E (post graduate) level. They encourage me for my starting journey of PHD to decide my research area and problem statement. Also I am thankful to my family members and my friends for their help and kindful support.

## REFERENCES

[1] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules in Large Databases,” Proc. 20th Int’l Conf. Very Large Data Bases, pp. 487-499, 1994.

[2] T. Calders and S. Verwer, “Three Naive Bayes Approaches for Discrimination-Free Classification,” Data Mining and Knowledge Discovery, vol. 21, no. 2, pp. 277-292, 2010.

[3] Sara Hajian and Josep Domingo-Ferrer, Fellow, IEEE, “A Methodology for Direct and Indirect Discrimination Prevention in Data Mining”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 7, JULY 2013, Published by the IEEE Computer Society

[4] Hongwei Mo and Lifang Xu, Automation College, Harbin Engineering University, “Immune Clone Algorithm for Mining Association Rules on Dynamic Databases”, Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI’05)

[5] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, “Discrimination Prevention in Data Mining for Intrusion and Crime Detection,” Proc. IEEE Symp. Computational Intelligence in Cyber Security (CICS ’11), pp. 47-54, 2011.

[6] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, “Rule Protection for Indirect Discrimination Prevention in Data Mining,” Proc. Eighth Int’l Conf. Modeling Decisions for Artificial Intelligence (MDAI ’11), pp. 211-222, 2011.

[7] F. Kamiran and T. Calders, “Classification without Discrimination,” Proc. IEEE Second Int’l Conf. Computer, Control and Comm. (IC4 ’09), 2009.

[8] F. Kamiran and T. Calders, “Classification with no Discrimination by Preferential Sampling,” Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.

[9] F. Kamiran, T. Calders, and M. Pechenizkiy, “Discrimination Aware Decision Tree Learning,” Proc. IEEE Int’l Conf. Data Mining (ICDM ’10), pp. 869-874, 2010.

[10] Chih-Chia Weng, Shan-Tai Chen, Hung-Che Lo Dept. of Computer Science, Chung Cheng Institute of Technology, National Defense University, Taiwan, R.O.C., “A Novel Algorithm for Completely Hiding Sensitive Association Rules”, Eighth International Conference on Intelligent Systems Design and Applications

[11] U.S. Federal Legislation. http://www.usdoj.gov

[12] D. Pedreschi, S. Ruggieri, and F. Turini, “Discrimination-Aware Data Mining,” Proc. 14th ACM Int’l Conf. Knowledge Discovery and Data Mining (KDD ’08), pp. 560-568, 2008.

[13] D. Pedreschi, S. Ruggieri, and F. Turini, “Measuring Discrimination in Socially-Sensitive Decision Records,” Proc. Ninth SIAM Data Mining Conf. (SDM ’09), pp. 581-592, 2009.

[14] D. Pedreschi, S. Ruggieri, and F. Turini, “Integrating Induction and Deduction for Finding Evidence of Discrimination,” Proc. 12th ACM Int’l Conf. Artificial Intelligence and Law (ICAIL ’09), pp. 157- 166, 2009.

[15] S. Ruggieri, D. Pedreschi, and F. Turini, “Data Mining for Discrimination Discovery,” ACM Trans. Knowledge Discovery from Data, vol. 4, no. 2, article 9, 2010.

[16] S. Ruggieri, D. Pedreschi, and F. Turini, “DCUBE: Discrimination Discovery in Databases,” Proc. ACM Int’l Conf. Management of Data (SIGMOD ’10), pp. 1127-1130, 2010.

[17] Faisal Kamiran, Toon Calders, Faculty of Mathematics and Computer Science, “Classification with No Discrimination by Preferential Sampling”, Eindhoven University of Technology The Netherlands

[18] Salvatore Ruggieri, Dino Pedreschi, Franco Turini, “Data Mining for Discrimination Discovery”, ACM Journal Name, Vol. V.

[19] R.Natarajan, Dr.R.Sugumar, M.Mahendran, K.Anbazhagan, “Design and Implement an Association Rule hiding Algorithm for Privacy Preserving Data Mining”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 7, September 2012

[20] “How to present discrimination claim” Handbook on non-discriminative directives

[21] C. Aggarwal and P. Yu. “Privacy-Preserving Data Mining: Models and Algorithms” , Springer, 2008.

[22] Faisal Kamiran, Toon Calders and Mykola Pechenizkiy, “Discrimination Aware Decision Tree Learning”, IEEE International Conference on Data Mining. IEEE press, 2010.