Abstract: Data mining is a more and more necessary technology for extracting useful data hidden in giant collections of information. There are, however, negative social perceptions concerning data processing, among that potential privacy invasion and potential discrimination. The latter consists of below the belt treating individuals on the premise of their happiness to a selected cluster. Automatic knowledge collection and data processing techniques like classification rule mining have paved the thanks to creating automatic selections, like loan granting/denial, premium computation, etc. If the coaching knowledge sets square measure biased in what regards discriminatory (sensitive) attributes like gender, race, religion, etc., discriminatory selections might result. For this reason, antidiscrimination techniques together with discrimination discovery and interference are introduced in data processing. Discrimination is often either direct or indirect. Direct discrimination happens once selections square measure created supported sensitive attributes. Indirect discrimination happens once selections square measure created supported no sensitive attributes that square measure powerfully correlative with biased sensitive ones. During this paper, we have a tendency to tackle discrimination interference in data processing and propose new techniques applicable for direct or indirect discrimination interference severally or each at an equivalent time. We have a tendency to also propose new metrics to gauge the utility of the planned approaches and that we compare these approaches. The experimental evaluations demonstrate that the planned techniques square measure effective at removing direct and/or indirect discrimination biases within the original knowledge set whereas protective knowledge quality

I. INTRODUCTION

In social science, it involves denying to members of 1 cluster opportunities that are out there to different teams. there’s an inventory of antidiscrimination acts, that are laws designed to forestall discrimination on the idea of variety of attributes (e.g., race, religion, gender, position, disability, legal status, and age) in numerous settings (e.g., employment and coaching, access to public services, credit and insurance, etc.). As an example, the ecru Union implements the principle of equal treatment between men and girls within the access to and provider of products and services in [3] or in matters of employment and occupation in [4]. Though there are some laws against discrimination, all of them are reactive, not proactive. Technology will add proactively to legislation by contributory discrimination discovery and bar techniques.

Discrimination is either direct or indirect (also known as systematic). Direct discrimination consists of rules or procedures that expressly mention minority or underprivileged teams supported sensitive discriminatory attributes associated with cluster membership. Indirect discrimination consists of rules or procedures that, whereas not expressly mentioning discriminatory attributes, on purpose or accidentally might generate discriminatory selections. Redlining by money establishments (refusing to grant mortgages or insurances in urban areas they think about as deteriorating) is AN prototypical example of indirect discrimination, though in no way the sole one. With a small abuse of language for the sake of compactness, into his paper indirect discrimination also will be noted as redlining and rules inflicting indirect discrimination are going to be known as redlining rules [12]. Indirect discrimination might happen thanks to the supply of some background (rules), as an example that an explicit postal code corresponds to a deteriorating space or a neighborhood with largely black population. The background could be accessible from in public out there knowledge (e.g., census knowledge) or could be obtained from the initial knowledge set itself thanks to the existence of nondiscriminatory attributes that are extremely correlate with the sensitive ones within the original data set our study.

Indirect discrimination might happen thanks to the supply of some background (rules).
II. RELATED WORK

Preprocessing. 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.

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 non-discriminatory constraint is embedded into a decision tree learner by changing its splitting criterion and pruning strategy through a novel leaf relabeling approach.

III. A PROPOSAL FOR DIRECT AND INDIRECT

A. The Approach

Our approach for direct and indirect discrimination prevention can be described in terms of two phases:

- Discrimination measurement. Direct and indirect-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. Second, direct discrimination is measured by identifying _-discriminatory rules among the PD rules using a direct discrimination measure (elite) and a discriminatory threshold (_). Third, indirect discrimination is measured by identifying redlining rules among the PND rules combined with background knowledge, using an indirect discriminatory measure (elb), and a discriminatory threshold (_). Let MR be the database of direct _-discriminatory rules obtained with the above process. In addition, let RR be the database of redlining rules and their respective indirect _-discriminatory rules obtained with the above process. Discrimination discovery includes identifying

- 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.

B. Data Transformation for Direct Discrimination

In order to convert each _-discriminatory rule into an _-protective rule, based on the direct discriminatory measure (i.e., Definition 2), we should enforce the following inequality for each _-discriminatory rule r0 : A.B -> C in MR, where A is a discriminatory item set: that if each _-discriminatory item set would be free of direct discrimination.

IV. BACKGROUND

A. Basic Definitions

- A data set is a collection of data objects (records) and their attributes. Let DB be the original data set. An item is an attribute along with its value, e.g., Race = black.

- An item set, i.e., X, is a collection of one or more items, e.g., {Foreign worker =Yes, City = NYC}.

- A classification rule is an expression X --> C, where C is a class item (a yes/no decision), and X is an item set containing no class item, e.g., {Foreign worker = Yes; City = NYC} --> Hire = no. X is called the premise of the rule.

- The support of an item set, supp(X), is the fraction of records that contain the item set X. We say that a rule X --> C is completely supported by a record if both X and C appear in the record.

- The confidence of a classification rule, conf(X --> C), measures how often the class item C appears in records that contain X. Hence, if supp(X) = 0 the

\[
\text{Con}(X --> C) = \frac{\text{supp}(X, C)}{\text{supp}(X)}
\]

\[
\text{conf}(X \rightarrow C) = \frac{\text{supp}(X, C)}{\text{supp}(X)}.
\]

V. ALGORITHMS

We describe in this section our algorithms based on the direct and indirect discrimination prevention methods proposed in Sections 3.2, 3.3, and 3.4. There are some assumptions common to all algorithms in this section. First, we assume the class attribute in the original data set DB to be binary (e.g., denying or granting credit). Second, we consider classification rules with ergative decision (e.g., denying credit) to be in FR. Third, we assume the discriminatory item sets (i.e., A) and the non-discriminatory item sets (i.e., D) to be binary or no binary categorical.

Direct discrimination prevention algorithms, we start with direct rule protection. Algorithm 1 details Method 1 for DRP. For each direct _-discriminatory rule r0 in MR (Step 3), after finding the subset DBc (Step 5), records in DBc should be changed until the direct rule protection requirement (Step 10) is met for each respective rule (Steps 10-14).
Algorithm 1. DIRECT RULE PROTECTION (METHOD 1)
1: Inputs: DB, FR, MR, α, DI
2: Output: DB' (transformed data set)
3: for each r' : A, B → C ∈ MR do
4:    FR ← FR − {r'}
5:    DBc ← All records completely supporting ¬A, B → ¬C
6:    for each dbc ∈ DBc do
7:        Compute impact(dbc) = |{ra ∈ FR|dbc supports the premise of ra}|
8:    end for
9:    Sort DBc by ascending impact
10:   while conf(r') ≥ α · conf(B → C) do
11:      Select first record in DBc
12:      Modify discriminatory item set of dbc from ¬A to A in DB
13:      Recompute conf(r')
14:    end while
15: end for
16: Output: DB' = DB

Algorithm 2. DIRECT RULE PROTECTION (METHOD 2)
1: Inputs: DB, FR, MR, α, DI
2: Output: DB' (transformed data set)
3: for each r' : A, B → C ∈ MR do
4:    Steps 4-9 Algorithm 1
5:    while conf(B → C) ≤ \(\frac{\text{conf}(r')}{\alpha}\) do
6:        Select first record in DBc
7:        Modify the class item of dbc from ¬C to C in DB
8:        Recompute conf(B → C)
9:    end while
10: end for
11: Output: DB' = DB

Algorithm 3. DIRECT RULE PROTECTION AND RULE GENERALIZATION
1: Inputs: DB, FR, TR, p ≥ 0.8, α, DI
2: Output: DB' (transformed data set)
3: for each r' : A, B → C ∈ TR do
4:    FR ← FR − {r'}
5:    if TRc = RG then
6:        // Rule Generalization
7:        DBc ← All records completely supporting A, B, ¬D → C
8:        Steps 6-9 Algorithm 1
9:    while conf(r') > \(\frac{\text{conf}(r')}{p \cdot \text{conf}(D, B, ¬C)}\) do
10:       Select first record in DBc
11:      Modify class item of dbc from C to ¬C in DB
12:      Recompute conf(r')
13:    end while
14:    end if
15:    if TRc = DRP then
16:        // Direct Rule Protection
17:        Steps 5-14 Algorithm 1 or Steps 4-9 Algorithm 2
18:    end if
19: end for
20: Output: DB' = DB
VI. CONCLUSION AND FUTURE WORK

It is more than obvious that most people do not want to be discriminated because of their gender, religion, nationality, age, and so on, especially when those attributes are used for making decisions about them like giving them a job, loan, insurance, etc. The purpose of this paper was to develop a new pre-processing discrimination prevention methodology including different data transformation methods that can prevent direct discrimination, indirect discrimination or both of them at the same time.
REFERENCES


