International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-4 Issue-6, August 2015

An Enhanced Approach for Privacy Preservation in Anti-Discrimination Techniques of Data Mining

Sreejith S., Sujitha S.

Abstract*— Data mining is an important area for extracting useful information from large collections of data. There are mainly two threats for individuals whose information is published: privacy and discrimination. Privacy invasion occurs when the values of published sensitive attributes is linked to specific individuals. Discrimination is the unfair or unequal treatment of people based on their membership to a specific category, group or minority. In data mining, decision models are mainly derived on the basis of records stored by means of various data mining methods. But there may be a risk that the extracted knowledge may impose discrimination. Many organizations collect a lot of data also for decision making. The sensitive information of the individual whom the published data relate to, may be revealed, if the data owner publishes the data directly. Hence, discrimination prevention and privacy preservation need to be ensured simultaneously in the decision making process. In this paper, discrimination prevention along with different privacy protection techniques have been proposed and the utility measures have been evaluated.

Index Terms*— Discriminatory attribute, direct discrimination prevention, indirect discrimination prevention, rule generalization, rule protection, k-anonymity, l- diversity, t-closeness

I. INTRODUCTION

Data mining is the process of discovering useful knowledge or patterns from large datasets. While extracting hidden information, the process may impose the risk of violation of non-discrimination and privacy in the dataset. Privacy refers to the individual right to choose freely what to do with one's own personal information whereas discrimination refers to unfair or unequal treatment of people based on membership to a category or a group.

Discrimination denies opportunities, for members of one group, which are available to other groups. In the data mining process, if the training dataset itself are biased for or against a particular community, then the decisions may also show discriminatory behavior. Therefore, to discover and eliminate such biases from the data, without harming the decision making utility, is very important and crucial.

Discrimination in the dataset are of two types: direct discrimination and indirect discrimination. Direct discrimination are the rules or procedures that explicitly mention separate groups based on the discriminatory attributes whereas indirect discrimination are the rules or procedures that do not explicitly mention the discriminatory attributes but unintentionally generates discriminatory attributes. When organizations such as banks, educational institutions, and insurance companies offer many services making use of user's data, the sensitive information of the individual may get revealed. Therefore, the data need to be processed to protect against privacy intrusion, while the utility of the data is preserved as much as possible. This is the privacy-preserving data publishing (PPDP). Consider the case where a set of patterns are extracted from personal data of a population and is used for making useful decisions. These set of patterns may reveal sensitive information about individual persons in the training population. The decision rules based on such patterns may also sometimes lead to discrimination. Therefore, it is highly essential to protect the dataset from privacy intrusion and discriminatory decisions.

This paper is organized as follows: Section 2 discusses the existing work dealing with anti-discrimination and privacy preservation in data mining. Section 3 deals with the background information, basic definitions of discrimination prevention and privacy protection in data mining. Section 4 describes the method for simultaneous direct and indirect discrimination prevention and Section 5 proposes a method for anti-discrimination along with privacy models of *k*-anonymity and *l*-diversity. Section 6 gives the experimental details and results of the proposed method and Section 7 concludes this paper on its work.

II. RELATED WORK

In this section, the existing work dealing with antidiscrimination and privacy protection in data mining is discussed.

D. Pedreschi, S. Ruggieri and F. Turini (2008) presented the first paper which addresses the discrimination problem in data mining models [1]. They investigated how discrimination is hidden in data mining models and measured discrimination through the introduction of lift. They also introduced α protection as a measure of the discrimination and proposed the extraction of classification rules. F. Kamiran and T. Calders (2009) tackled the problem of classification scheme for learning unbiased models on biased training data [2]. The method is based on massaging the dataset by making the least modifications that leads to an unbiased dataset. But the main drawback was that numerical attributes and group of attributes were not considered as sensitive attribute. D. Pedreschi, S. Ruggieri and F. Turini (2009) presented a systematic framework for measuring discrimination, based on the analysis of decision records [3]. They investigated whether direct and indirect discrimination can be found in a given set of records by measuring the degree of discrimination.

Manuscript published on 30 August 2015.
* Correspondence Author (s)
Sreejith S., Department of Computer Science & Engineering, L B S Institute of Technology for Women, Thiruvananthapuram Kerala. India.
Sujitha S., Department of Computer Science & Engineering, L B S Institute of Technology for Women, Thiruvananthapuram Kerala. India.

They discussed integrating induction, through classification rule extraction, and deduction through a computational logic implementation of the analytical tools. In 2010, they also presented the discrimination discovery in databases in which unfair practices are hidden in a dataset of historical decisions [4]. The DCUBE system was based on the classification rule extraction and it also briefs the legal issues about discrimination hidden in the data.

T. Calders and S. Verwer (2010) presented a modified Naive- Bayes classification approach that performs classification of the data, focussing on independent sensitive attribute. But the approach did not consider numerical attribute [5]. Faisal Kamiran, Toon Calders and Mykola Pechenizkiy (2010) also introduced two different techniques for incorporating discrimination measures into decision tree construction process [6].

S. Hajian, J. D. Ferrer and A. Martinez-Balleste (2011) introduced anti-discrimination in the context of cyber security [7]. They proposed data transformation method for discrimination prevention and considered several discriminatory attributes and their combinations. The issue of data quality was also addressed in the paper.

I. Zliobaitye, F. Kamiran and T. Calders (2011) studied how to train classifiers on historical data, used for supervised learning, so that they are discrimination free with respect to a given sensitive attribute [8]. They also analysed and introduced the conditional non-discrimination in classifier design and developed local techniques for handling conditional discrimination.

In 2012, F. Kamiran and T. Calders [9] presented the algorithmic solutions to preprocess the data removing discrimination before a classifier is learned.

Some studies showed that [10-12], it is not enough that just removing the identifier from the dataset protects it from privacy invasion. To reduce the probability that the attacker can retrieve the corresponding record of the victim from the background information, k-anonymity was first presented by Samarati and Sweeney [13]. To address the attacks based on lack of diversity in sensitive attribute values, l-diversity [14] was introduced.

A table satisfies p-sensitive k-anonymity property if it satisfies k-anonymity, and for each equivalent class, the number of distinct values for each sensitive attribute is at least p. (α, k)-anonymity [15], which combines k-anonymity and l-diversity, for each equivalent class, there are at least k records and the frequency of each sensitive attribute value does not exceed a threshold α. An equivalent class E is said to have t-closeness, if the semantic distance between the two distributions of the sensitive attribute values in E and data table is no more than a threshold t [16].

III. BACKGROUND

In this section, some of the background details required for the anti-discrimination and privacy preservation process is discussed.

A. Basic Definitions on Discrimination Prevention

Some of the basic definitions related to discrimination prevention data mining [19] are discussed below:

- A *data set* is a collection of data objects (records) and their attributes.
- 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 $X \rightarrow C$, 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.
- The *support* of an item set, *supp(X)*, is the fraction of records that contain the item set X. A rule $X \rightarrow C$ is completely supported by a record if both X and C appear in the record.
- The *confidence* of a classification rule, *conf*$(X \rightarrow C)$, measures how often the class item C appears in records that contain X.
- The negated item set, i.e., ¬X is an item set with the same attributes as X, but the attributes in ¬ X take any value except those taken by attributes in X.

B. Potentially Discriminatory and Nondiscriminatory Classification Rules

Let DI_s be the set of predetermined discriminatory items in *DB* (e.g., DIs = {Foreign worker = Yes; Race = Black; Gender = Female}).

1. A classification rule $X \rightarrow C$ is potentially discriminatory (PD) when $X = A, B$ with $A \subseteq DI_s$ a nonempty discriminatory item set and *B* a non-discriminatory item set. For e.g.: {Foreign worker = Yes; City = NYC} → Hire = No.
2. A classification rule $X \rightarrow C$ is potentially non-discriminatory (PND) when $X = D, B$ is a non-discriminatory item set. For e.g.: {Zip = 10451; City = NYC} → Hire = No, or {Experience = Low; City = NYC} → Hire = No.

Pedreschi et al. [1] translated the qualitative statements into quantitative formal counterparts over classification rules and they introduced a family of measures of the degree of discrimination of a PD rule.

C. Direct Discrimination Measure

Direct discrimination can be measured and evaluated on the basis of the extended lift (elift) [19].

Definition 1: *Let* $A, B \rightarrow C$, *a classification rule such that* $conf(B \rightarrow C) > 0$. *The extended lift of the rule is*

$$elift(A, B \rightarrow C) = \frac{conf(A,B \rightarrow C)}{conf(B \rightarrow C)} \qquad (1)$$

Definition 2: *Let* $\alpha \in R$ *be a fixed threshold and let A be a discriminatory item set. A PD classification rule* $c = A, B \rightarrow C$ *is* α*-protective w.r.t. elift if elift(c)* $< \alpha$. *Otherwise, c is* α *-discriminatory.*

Direct discrimination discovery identifies α-discriminatory rules and these biased rules inferred directly from the discriminatory attributes are called direct α-discriminatory rules.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

D. Indirect Discrimination Measure

Indirect discrimination discovery identifies redlining rules which are indirectly inferred from non-discriminatory items because of their correlation with discriminatory ones. To determine the redlining rules. The theorem [19] stated below gives a lower bound for α-discrimination of PD classification rules, given information available in PND rules (Υ, δ), and information available from background rules (β_1, β_2). It is assumed that the background knowledge takes the form of classification rules relating a non-discriminatory item set D to a discriminatory item set A within the context B.

Theorem 1: *Let r: D, B → C be a PND classification rule, and let*

Υ= *conf (r: D, B → C)* δ = *conf(B→ C)> 0*

Let A be a discriminatory item set, and let β1, β2 such that

conf $(r_{b1}: A, B \rightarrow D) \geq \beta_1$

conf $(r_{b2}: D, B \rightarrow A) \geq \beta_2 > 0$.

Call

$f(x)=\beta_1/\beta_2(\beta_2+x-1)$

$$elb(x,y) = \begin{cases} \dfrac{f(x)}{y} \; if \; f(x) > 0 \\ 0 \quad otherwise \end{cases}$$

It holds that, for $\alpha \geq 0$, *if elb* $(\Upsilon,\delta) \geq \alpha$, *the PD classification rule r': A, B → C is α- discriminatory.*

Based on this, redlining and nonredlining rules are stated as below:

Definition 3: A PND classification rule r: D, B → C is a redlining rule if it could yield an α-discriminatory rule r': A, B → C in combination with currently available background knowledge rules of the form r_{b1}: A, B → D and r_{b2}: D, B → A, where A is a discriminatory item set.

Definition 4: A PND classification rule r: D, B → C is a nonredlining or legitimate rule if it cannot yield any α-discriminatory rule r': A, B → C in combination with currently available background knowledge rules of the form r_{b1} : A, B → D and r_{b2} : D, B → A, where A is a discriminatory item set.

E. Basic Definitions on Privacy Preservation

The background knowledge required for reviewing privacy preservation techniques are discussed below:

Given the data table D (A_1… An), a set of attributes A = {A_1….A_n}, and a record/tuple t ∈ D.

- T [A_i….A_j]: sequence of the values A_i…A_j in t where {A_i….A_j} ⊆ {A_1….A_n}.
- D [A_i….A_j]: the projection maintaining duplicate records of attributes A_i….A_j in D.
- |D|: the cardinality of D.
- Identifiers are attributes that uniquely identify individuals in the database, like Passport number.
- A quasi-identifier (QI) is a set of attributes that, in combination, can be linked to external identified information for re-identifying an individual, for example: Zip code, Birthdate and Gender.
- Sensitive attributes (S) are those that contain sensitive information, such as Disease or Salary. Let S be a set of sensitive attributes in D.

F. Privacy Models

Definition 5: *k*- anonymity [13]

Let $D(A_1,.....,A_n)$, *be a data table and* $QI = \{Q_1,...,Q_m\} \subseteq (A_1,.....,A_n)$, *be a quasi-identifier. D is said to satisfy k-anonymity w.r.t. QI if each combination of values of attributes in QI is shared by at least k tuples (records) in D.*

The table that satisfies *k*- anonymity is said to be a *k*-anonymous table and the set of records which are similar on the QI values is referred to an equivalence class (EC). The probability of identifying an individual is reduced to 1/*k*. A larger *k* can bring a lower probability of a linkage attack. *k*-anonymity can be achieved by QI generalization or QI suppression. For example, Table 1(a) is a raw data and Table 1(b) is its corresponding 3- anonymous table.

Table 1. The anonymization of a subset of a data table

Personal status and sex	Present Residence Since	Age	Credit history
'male single'	4	67	'critical/other existing credit'
'female div/dep/mar'	3	22	existing paid'
'male single'	3	49	'no credits'
'male single'	1	45	'delayed previously'
'male single'	2	53	'no credits/all paid'
'male single'	1	35	'all paid'

(a): Raw data

Personal status and sex	Present Residence Since	Age	Credit history
'male single'	3-4	67	'critical/other existing credit'
'female div/dep/mar'	3-4	22	existing paid'
'male single'	3-4	49	'no credits'
'male single'	1-2	45	'delayed previously'
'male single'	1-2	53	'no credits/all paid'
'male single'	1-2	35	'all paid'

(b): 3-anonymous data

Personal status and sex	Present Residence Since	Age	Credit history
'male single'	3-4	67	'critical/other existing credit'
'female div/dep/mar'	3-4	22	existing paid'
'male single'	3-4	49	'no credits'

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

'male single'	1-2	45	'delayed previously'
'male single'	1-2	53	'no credits/all paid'
'male single'	1-2	35	'all paid'

(c): 3- diverse data

The main attacks identified in *k*- anonymity model are: homogeneity attack and background knowledge attack. In the above example, records may show a generality within an EC. Therefore, it is necessary to avoid a generality on SA. Generality attack is the local property of an EC when the S has distinct values within an EC such that an attacker can get important information. Sensitivity attack is the generality attack on SA sensitivity level, when all S values belong to higher sensitivity level than the average and similarity attack is the generality attack on sensitive semantic category when all S valued are similar based on the category.

Definition 6: *l*-diversity [14]

A q-block is l-diverse if it contains at least well-represented values for the sensitive attribute S.*

Distinct l- diversity does not prevent probabilistic attacks. An equivalence class can have one value appear more frequently than other values which enables the attacker to infer that an entity in the equivalence class is more likely to have that value. *l*-diversity may be sometimes difficult and unnecessary to achieve and it is insufficient to prevent attribute disclosure. *l*- diversity principle can sometimes lead to skewness attack and similarity attack.

Definition 7: *t*- closeness [16]

An equivalence class is said to have t-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole table is no more than a threshold t.

G. *Sanitation Mechanisms*

The sanitation mechanism implemented through anonymization was proposed by Samarati and Sweeney [13, 17, 18]. The computational procedures are based on generalization and suppression.

Generalization is the process of replacing the QI attribute values with a generalized version of them using the generalization taxonomy tree. There are five possible generalization mechanisms.

In full-domain generalization, all values in an attribute are generalized to the same level of the taxonomy tree. In subtree generalization, either all child values or none are generalized (at a nonleaf node). Sibling generalization is similar to the subtree generalization, except that some siblings may remain ungeneralised. In cell generalization, some instances of a value may remain ungeneralised while other instances are generalized. Multidimensional generalization allows two QI groups, even having the same value, to be independently generalized into different parent groups.

Suppression is the process of suppressing some values of the QI attributes for some (or all) records. Global recoding is the method of generalizing an attribute in Parent Table, through all the tuples, to the same level in the respective generalization hierarchy of that attribute. Local recoding is also referred to as cell-level generalization.

IV. Direct and Indirect Discrimination Prevention

A. *Discrimination Measurement*

Direct and indirect discrimination discovery is the process of identifying α-discriminatory rules and redlining rules. First, based on the predetermined discriminatory items in the dataset, frequent classification rules are divided into two groups: PD and PND rules. Direct discrimination can be found by identifying α-discriminatory rules among the PD rules using a direct discrimination measure and a discriminatory threshold. Indirect discrimination is measured by identifying redlining rules among the PND rules combined with background knowledge. Let MR be the database of direct α-discriminatory rules and RR be the database of redlining rules and their respective indirect α-discriminatory rules. Next, the original data is transformed for each respective α-discriminatory rule, without affecting the data or other rules and is evaluated to check whether they are free of discrimination.

B. *Data Transformation for Direct Discrimination*

The approach of data transformation for preventing direct discrimination is based on the fact that the data set of decision rules would be free of direct discrimination if it only contains PD rules that are α-protective or are instances of at least one nonredlining PND rule. So the transformation methods should be applied in such a way that minimum information loss occurs and each α- discriminatory rule either becomes α-protective or an instance of a nonredlining rule. Direct rule protection [19] implements two methods of data transformation

1. *Direct Rule Protection*

To convert each α-discriminatory rule in *MR* into an α-protective rule, the inequality r': A, B → C is implemented for each α- discriminatory rule in *MR*.

$$elift(r') < \alpha \qquad (2)$$

By using the statement of the *elift* Definition, Inequality (2) can be rewritten as

$$\frac{conf(r': A, B \rightarrow C)}{conf(B \rightarrow C)} < \alpha \qquad (3)$$

The above inequality can also be written in the following way:

$$conf(r': A, B \rightarrow C < \alpha.conf(B \rightarrow C) \qquad (4)$$

So, Inequality (2) can be achieved by decreasing the confidence of the α-discriminatory rule *r': A, B → C* to a value less than *α.conf (B→C)*. A solution for decreasing *conf(r': A, B → C)* is to alter the discriminatory item set from ¬*A* to *A* in the subset DB_c of all records which completely support the rule ¬*A*, *B* → ¬*C* so that minimum impact occurs on other rules as,

$$conf(r': A, B \rightarrow C) = \frac{supp(A, B, C)}{supp(A, B)} \qquad (5)$$

Inequality (3) can also be written as,

$$conf(B \rightarrow C) > \frac{conf(r': A, B \rightarrow C)}{\alpha} \qquad (6)$$

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

So, Inequality (2) can also be achieved by increasing the confidence of the rule $(B \rightarrow C)$ of the α-discriminatory rule *r': A, B→C* to a value higher than the right-hand side of Inequality (6). It can be implemented by altering the class item from $\neg C$ to C in the subset DB_c of all records which completely support the rule $\neg A, B \rightarrow \neg C$ so that minimum impact occurs on other rules as,

$$conf(B \rightarrow C) = \frac{supp(B,C)}{supp(B)} \quad (7)$$

Therefore, in direct rule protection, Method 1 changes the discriminatory item set in some records, while Method 2 changes the class item in some records.

2. Rule Generalization

Rule Generalization method is based on the fact that if each α-discriminatory rule r' : A,B → C in the database of decision rules was an instance of at least one nonredlining PND rule r : D, B → C, the data set would be free of direct discrimination.

Definition 6: *Let p ∈[0, 1]. A classification rule r': A, B → C is a p-instance of r: D, B → C if both conditions below are true:*

- *Condition 1: conf(r) ≥ p. conf(r')*
- *Condition 2: conf(r'': A, B → D) ≥ p.*

Rule generalization can be applied only for *α*-discriminatory rules *r'* for which there is at least one nonredlining PND rule *r* satisfying at least one of the two conditions of Definition 6. If any of the two conditions does not work, the original dataset needs to be transformed Assume that Condition 2 is satisfied and Condition 1 is not satisfied. Based on the definition 6,

$$conf(r: D, B \rightarrow C) \geq p.\, conf(r': A, B \rightarrow C) \quad (8)$$

or

$$conf(r': A, B \rightarrow C) \leq \frac{conf(r: D, B \rightarrow C)}{p} \quad (9)$$

So, Inequality (8) can be obtained by decreasing the confidence of the *α*-discriminatory rule *(r': A, B → C)* to values less than the right-hand side of Inequality (9), without affecting the confidence of rule *r: D, B → C* or the satisfaction of Condition 2 of Definition 6. A suitable way to decrease this confidence is to alter the class item from *C* to $\neg C$ in the subset DB_c of all records in the original data set which completely support the rule A, B, ¬D → C and have minimum impact on other rules.

Now, assume that Condition 1 of definition 6 is satisfied and Condition 2 is not satisfied. i.e. ,

$$conf(r'': A, B \rightarrow D) \geq p \quad (10)$$

The above inequality (Inequality (11)) can be satisfied by increasing the confidence of rule *r'': A, B → D* to a value higher than p, without affecting the satisfaction of Condition 1.But, increasing the confidence of *r''* impacts on the confidence of the *r* or *r'* and can affect the satisfaction of Condition 1 of Definition 5; Hence to increase the confidence of *r'', supp(A,B)* is decreased (which increases *conf(r')*) or ¬D is altered to D for those records satisfying *A* and *B* (which decreases *conf(r)*). Therefore, rule generalization can only be applied if Condition 2 is satisfied without any data transformation.

3. Direct Discrimination Prevention Algorithms

In this section, the algorithms used for direct discrimination prevention was proposed [19]. There are some assumptions applied to these algorithms. First, the class attribute in the original data set *DB* is assumed to be a binary value. Second, the classification rules with negative decisions are in FR are only considered. Third, we assume the discriminatory item sets and the non-discriminatory item sets is assumed to have a binary value.

Algorithm 1: Direct Rule Protection (Method 1)

Step 1: Inputs: *DB, FR, MR,α, DI$_s$*
Step 2: Output: DB' (transformed data set)
Step 3: for each *r': A, B → C ∈ MR* do
Step 4: *FR ←FR- {r'}*
Step 5: *DB$_c$* ←All records completely supporting *¬A, B→¬C*
Step 6: for each db$_c$ ∈ *DBc* do
Step 7: Compute *impact (db$_c$)* = *|{r$_a$ ∈FR|db$_c$* supports the premise of *ra}|*
Step 8: end for
Step 9: Sort *DB$_c$* by ascending impact
Step 10: while *conf (r')≥ conf (B → C)* do
Step 11: Select first record in *DB$_c$*
Step 12: Modify discriminatory item set of *db$_c$* from *¬A to A* in *DB*
Step 13: Recompute *conf (r')*
Step 14: end while
Step 15: end for
Step 16: Output: *DB'=DB*

Algorithm 1 gives the implementation of direct rule protection using Method 1. For each direct *α*-discriminatory rule *r'* in *MR* (Step 3), we find a subset DBc, the records of which we find the impact of those rules on other *α*- protective rules. The lowest impact rules are used for transformation (Step 10-13). All the above process are repeated for each *α*-discriminatory rule in *MR*.

Algorithm 2: Direct Rule Protection (Method 2)

Step 1: Inputs: *DB, FR, MR,α, DI$_s$*
Step 2: Output: DB' (transformed data set)
Step 3: for each *r: A, B → C ∈ MR* do
Step 4: Steps 4-9 Algorithm1
Step 5: while *conf (B→C)≤ conf (r')/ α* do
Step 6: Select first record in *DB$_c$*
Step 7: Modify discriminatory item set of *db$_c$* from *¬C to C* in *DB*
Step 8: Recompute *conf (B→C')*
Step 9: end while
Step 10: end for
Step 11: Output: *DB'=DB*

The above algorithm is implemented for Method 2 of DRP. The subset finding and impact minimization steps are same as that of Algorithm 1. The class item of the discriminatory item set is modified here instead of the discriminatory attribute in DRP Method 1.

Algorithm 3: Direct Rule Protection and Rule Generalization

Step 1: Inputs: *DB, FR, TR,α, p≥ 0.8, DI$_s$*
Step 2: Output: DB' (transformed data set)

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Step 3: for each *r': A, B → C ∈ TR* do
Step 4: *FR ←FR- {r'}*
Step 5: if $TR_{r'}$ = *RG* then
Step 6: // Rule Generalization
Step 7: DB_c ←All records completely supporting A, B, ¬D → C
Step 8: Steps 6-9 Algorithm 1
Step 9: while *conf (r') > conf* (r_b*:D, B→C)/p* do
Step 10: Select first record in DBc
Step 11: Modify class item of db_c from *C* to ¬*C* in *DB*
Step 12: Recompute *conf (r')*
Step 13: end while
Step 14: end if
Step 15: if $TR_{r'}$ = *DRP* then
Step 16: // Direct Rule Protection
Step 17: Steps 5-14 Algorithm 1 or Steps 4-9 Algorithm 2
Step 18: end if
Step 19: end for
Step 20: Output: *DB'=DB*

The above algorithm as input *TR*, which contains all r'∈ MR and their respective $TR_{r'}$ and r_b. For each *α*-discriminatory rule *r'* in *TR*, if *TRr'* shows that rule generalization should be performed (Step 5), after determining the records that should be changed for impact minimization, these records should be changed until the rule generalization requirement is met (Steps 9-13). Also, if $TR_{r'}$ shows that direct rule protection should be performed (Step 15), either Method 1 or Method 2, can be implemented (Step 17).

C. *Data Transformation for Indirect Discrimination*

Indirect discrimination prevention is based on the principle that data set of decision rules would be free of indirect discrimination if it contained no redlining rules. Indirect Rule Protection (IRP) converts the redlining rules to nonredlining rules.

1. Indirect Rule Protection

To convert a redlining rule into a nonredlining rule, the following inequality, for each r: D, B → C in RR, is implemented.

$$\text{elb}(\Upsilon,\delta)<\alpha \qquad (11)$$

The above inequality can also be written as, (using the definitions of elb from Theorem 1)

$$\frac{\frac{conf(rb1)}{conf(rb2)}conf(rb2)+conf(r)-1}{conf(B\rightarrow C)}<\alpha \qquad (12)$$

or

$$conf(rb1)<\frac{\alpha.conf(B\rightarrow C).conf(rb2)}{conf(rb2)+conf(r)-1} \qquad (13)$$

Hence, inequality (11) can be satisfied by decreasing the confidence of rule r_{b1}: A, B → D to values less than the right-hand side of Inequality (13) without affecting either the confidence of the redlining rule, r or the confidence of the B → C and r_{b2} rules. It can be done by to altering the discriminatory item set from ¬A to A in the subset DB_c of all records which completely support the rule ¬A, B, ¬D → ¬C so that minimum impact occurs on other rules as,

$$conf(A,B\rightarrow D=\frac{supp(A,B,D)}{supp(A,B)} \qquad (14)$$

Inequality (12) can also be written as

$$conf(B\rightarrow C)>\frac{\frac{conf(rb1)}{conf(rb2)}conf(rb2)+conf(r)-1}{\alpha} \qquad (15)$$

In this case, Inequality (11) can be achieved by increasing the confidence of the rule B → C of the redlining rule r: D, B → C to values greater than the right hand side of Inequality (15) without affecting either the confidence of the redlining rule, r or the confidence of the r_{b1} and r_{b2} rules. This can be done by altering the class item from ¬C to C in the subset DB_c of all records which completely support the rule ¬A, B, ¬D → ¬C so that minimum impact occurs on other rules.

Therefore, in indirect rule protection also, Method 1 changes the discriminatory item set in some records, while Method 2 changes the class item in some records.

2. Indirect Discrimination Prevention Algorithm

The algorithm stated below proposes the data transformation method for simultaneous direct and indirect discrimination prevention.

Algorithm 4: Direct and Indirect Discrimination Prevention

Step 1: Inputs: DB, FR, MR,α, DI_s
Step 2: Output: DB' (transformed data set)
Step 3: for each r: X → C ∈ RR, where D, B ⊆ X do
Step 4: ϒ = conf(r)
Step 5: for each r: (A ⊆ DI_s), (B ⊆ X) → C ∈ RR do
Step 6: β_2 = conf (r_{b2}: X → A)
Step 7: Δ_1 = supp (r_{b2}: X → A)
Step 8: δ = conf (B → C)
Step 9: Δ_2 = supp (B → A)
Step 10: $\beta_1 = \Delta_1/\Delta_2$ //conf (r_{b1}: A, B → D)
Step 11: Find DBc: all records in DB that completely support ¬A, B, ¬D → ¬C
Step 12: Steps 6-9 Algorithm 1
Step 13: if r' ∈ MR then
Step 14: while (δ≤(β_1(β_2+ϒ-1))/ (β_2*α)) and (δ ≤ (conf(r')/α)) do
Step 15: Select first record db_c in DBc
Step 16: Modify the class item of db_c from ¬C to C in DB
Step 17: Recompute δ = conf (B → C)
Step 18: end while
Step 19: else
Step 20: while (δ≤(β_1(β_2+ϒ-1) do
Step 21: Steps 15-17 Algorithm 4
Step 22: end while
Step 23: end if
Step 24: end for
Step 25: end for
Step 26: for each r': (A, B → C) ∈ MR\ RR do
Step 27: δ = conf (B → C)
Step 28: Find DB_c: all records in DB that completely support ¬A, B, → ¬C
Step 29: Step 12

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Step 30: while ($\delta \leq (conf(r')/\alpha)$) do
Step 31: Steps 15-17 Algorithm 4
Step 32: end while
Step 33: end for
Step 34: Output: DB'=DB

For the rules extracted from DB, if there is any overlap between direct and indirect discriminatory rules in MR and RR, data transformation is performed until both the direct and indirect rule protection conditions are satisfied. If there is no such overlap, data transformation method is adopted using the Method 2 for IRP until indirect discrimination prevention condition is satisfied. Then, for each direct discriminatory rule data transformation method is applied according to Method 2 for DRP.

V. A PROPOSAL FOR DISCRIMINATION PREVENTION AND PRIVACY PRESERVATION

During the data mining process, the set of data or patterns extracted from the original dataset can be used for the subsequent decision making scenarios. The set of patterns can sometimes reveal the sensitive information about individual persons and, some of the attributes or rules in the dataset may show discrimination against a particular community, which can lead to unfair discriminatory decisions. It is therefore highly essential and crucial that privacy and discrimination risks should be tackled.

In this section, discrimination discovery, its measurement and prevention methods are implemented along with different privacy models like *k*-anonymity and *l*- diversity. These methods are also evaluated using the utility measures and it is clear that *l*-diversity yields more efficient discrimination free dataset.

A. Discrimination Prevention and k- anonymity

To achieve *k*-anonymity, it required that each record should be indistinguishable with at least *k-1* records on QIs in anonymous table. From the original dataset, frequent classification rules are extracted using the Apriori algorithm [20]. For each of the frequent rules, anonymization can be applied for the quasi- attribute. *k*- anonymization can be applied either by QI generalization or QI suppression. The framework for achieving discrimination in the anonymized frequent rules in shown in the figure 1.

B. Discrimination Prevention and l- diversity

While *k*- anonymity protects the dataset against identity disclosure, it does not provide protection against attribute disclosure. The two main attacks in *k*- anonymity are homogeneity attack and background knowledge attack. From the homogenous distribution of the sensitive records and also from the background knowledge about the individual, an attacker can retrieve the sensitive information and it can also affect the discriminatory attribute. To address these limitations, *l*- diversity was introduced [14].

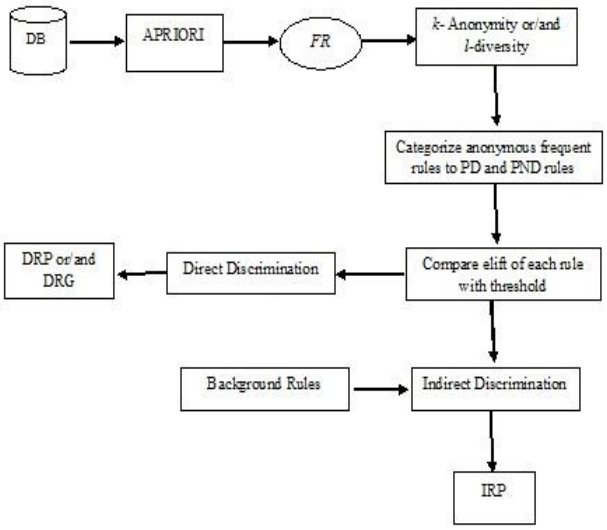

Figure 1: Discrimination Prevention and *k*- anonymity and/or *l*-diversity.

VI. EXPERIMENTS

This section details the experimental results of the direct and indirect discrimination methods using achieving *k*-anonymity and *l*- diversity. To obtain FR, the Apriori algorithm [20] which is used to extract frequent classification rules has been implemented.

A. Data Set

The dataset used here is the German credit data set [21] which consists of 1,000 records and 20 attributes (without class attribute) of bank account holders. The class attribute takes values representing good or bad classification. For the experiments with this data set, QI attribute is set as {Present residence since}, DI_s = {Foreign worker = Yes, Personal Status = Female and not Single, Age = Old}; (cut-off for Age = Old: 50 years old).

The Adult data set [22], also known as Census Income, is also used in the experiments. This data set consists of train part and test part. The data set has 14 attributes (without class attribute). The train part is used in the experiments. For the experiments with the Adult data set, we set DIs = Sex= Female, Age <=30 and QI=Age, Education, Marital Status.

B. Utility Measures

The discrimination prevention techniques can be evaluated on two basis: to measure the success in removing discrimination and to measure the impact of information loss [19].

- Discrimination Prevention Degree (DPD) is the percentage of α-discriminatory rules that are no longer α-discriminatory in the transformed data set.
- Discrimination Protection Preservation (DPP) is the percentage the α-protective rules in the original data set that remain α- protective in the transformed data set.
- Misses Cost (MC) is the percentage of rules extracted from the original dataset that transformed cannot be extracted from the transformed dataset.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

- Ghost Cost (GC) is the percentage of rules extracted from the transformed dataset that were not extracted from the original dataset.

C. Results and Discussion

Table 2 shows the results for direct and indirect discrimination prevention for minimum support 1 percent and minimum confidence 7 percent with German Credit Dataset. The privacy preservation techniques are *k*-anonymity and *l*- diversity. Tables 3 and 4 shows the utility measures after applying privacy preservation techniques. The results of the discrimination prevention methods are applied for the threshold value α= 1.0 and DI_s = {Foreign worker = Yes; Personal Status = Female and not Single; Age >=50}.Tables 4 and 5 shows the discrimination utility measures for privacy reservation techniques. From the below tables, it is shown that the discrimination prevention methods along with privacy preservation yields more efficient results. *l*-diversity is more useful than *k*- anonymity in the case where it is used for discrimination protection.

Methods	α	*p*	DPD	DPP	MC	GC
DRP (Method 1)	1.0	n. a	100	83.62	0	9.27
DRP (Method 2)	1.0	n. a	85.45	83.62	0	9.27
DRP+RG	1.0	0.8	n. a.	n. a.	n. a.	n. a.
DRP+IRP	1.0	n. a	84	83.62	0	9.27

Table 2: Discrimination Utility Measures (German Credit dataset)

Methods	α	*p*	DPD	DPP	MC	GC
DRP (Method 1)	1.0	n. a	92.21	100	15.44	13.52
DRP (Method 2)	1.0	n. a	93.80	100	0	4.6
DRP+RG	1.0	0.8	n. a.	n. a.	n. a.	n. a.
IRP (Method 1)	1.0	n. a	92.61	100	2.65	2.87
IRP (Method 2)	1.0	n. a	92.14	100	2.45	2.87
DRP+IRP	1.0	n. a	98.47	100	1.65	1.57

Table 3. Discrimination Utility Measures with *k*-anonymity (German Credit dataset)

Tables 5 shows discrimination utility measures incorporated with privacy preservation for Adult dataset with minimum confidence 7 percent and minimum support 1 percent. Tables 6 and 7 shows privacy preservation utility measures. It is clear in both the datasets that utility measures shows efficient results for privacy protection. The results of the discrimination prevention methods are applied for the threshold value α= 1.2 and DI_s = {Foreign worker = Yes; Personal Status = Female and not Single; Age >=50}.

Methods	α	*p*	DPD	DP	MC	GC
				P		
DRP (Method 1)	1.0	n. a	93.45	100	10.47	9.68
DRP (Method 2)	1.0	n. a	95.67	100	0	3.5
DRP + RG	1.0	0.8	n. a.	n. a.	n. a.	n. a.
IRP (Method 1)	1.0	n. a	94.87	100	2.62	2.85
IRP (Method 2)	1.0	n. a	93.56	100	2.42	2.56
DRP+ IRP	1.0	n. a	98.89	100	1.62	1.50

Table 4: Discrimination Utility Measures with *l*-diversity (German Credit dataset)

Methods	α	*p*	DPD	DPP	MC	GC
DRP (Method 1)	1.0	n. a	100	83.62	0	9.27
DRP (Method 2)	1.0	n. a	85.45	83.62	0	9.27
DRP+RG	1.0	0.8	n. a.	n. a.	n. a.	n. a.
DRP+IRP	1.0	n. a	84	83.62	0	9.27

Table 5: Discrimination Utility Measures (Adult dataset)

Methods	α	*p*	DPD	DPP	MC	GC
DRP (Method 1)	1.2	n. a	92.21	100	15.44	13.52
DRP (Method 2)	1.2	n. a	93.80	100	0	4.6
DRP+RG	1.2	0.8	n. a.	n. a.	n. a.	n. a.
IRP (Method 1)	1.2	n. a	92.61	100	2.65	2.87
IRP (Method 2)	1.2	n. a	92.14	100	2.45	2.87
DRP+IRP	1.2	n. a	98.47	100	1.65	1.57

Table 6. Discrimination Utility Measures with *k*-anonymity (Adult dataset)

Methods	α	*p*	DPD	DPP	MC	GC
DRP (Method 1)	1.0	n. a	93.45	100	10.47	9.68
DRP (Method 2)	1.0	n. a	95.67	100	0	3.5
DRP + RG	1.0	0.8	n. a.	n. a.	n. a.	n. a.
IRP (Method 1)	1.0	n. a	94.87	100	2.62	2.85
IRP (Method 2)	1.0	n. a	93.56	100	2.42	2.56
DRP+ IRP	1.0	n.a	98.89	100	1.62	1.50

Table 7. Discrimination Utility Measures with *l*-diversity (Adult dataset)

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Figure 2 shows the behaviour of information loss for different values of α = 1.0, 1.1 and 1.2 in the German credit dataset. It is clear that as the value of α increases, the number of discriminatory rules becomes less and the impact on α protective rules is minimum. So data transformation is needed only for less number of rules and hence information loss is minimum.

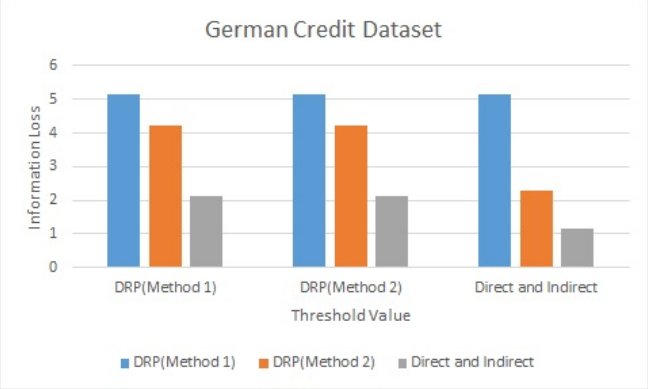

Figure 2: Information Loss for German Credit dataset

Figure 3 shows the behaviour of information loss for different values of α = 1.2, 1.3 and 1.4 in the Adult dataset. In this experiment also, as data transformation occurs for less number of rules if the value of α is increased and hence information loss is minimum.

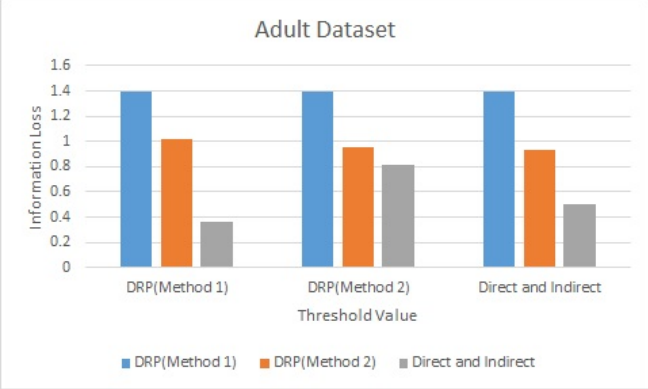

Figure 3: Information Loss for Adult dataset

VII. CONCLUSION

Data mining is an important technology for extracting useful knowledge hidden in large collections of data. Privacy preserving and anti-discrimination techniques have been introduced in data mining to protect the sensitive data before it is being published. In this paper, privacy protection is achieved by *k*- anonymity and *l*- diversity models. Direct and indirect discrimination prevention is implemented by rule protection and rule generalization methods and evaluated using utility measures. Discrimination free and privacy protected frequent patterns sets are also extracted in this paper. It is shown that *l*-diversity can be used to achieve more efficient privacy protection and discrimination free datasets than *k*- anonymity. In the future, these data transformation methods to prevent discrimination can be applied to other data anonymization techniques. The existing approaches can also be used for other data mining tasks.

ACKNOWLEDGMENT

In this paper I would like to thank the Dept. of CSE, LBSITW, for giving me an opportunity to write this paper and also to all the authors of the papers which I refer, to get enough information to write this paper.

REFERENCES

1. 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.
2. F. Kamiran and T. Calders, “Classification without Discrimination,” Proc. IEEE Second Int’l Conf. Computer, Control and Comm. (IC4 ’09), 2009.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. F. Kamiran and T. Calders, “Classification without Discrimination,” Proc. IEEE Second Int’l Conf. Computer, Control and Comm. (IC 4 ’09), 2009.
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. Sweeney L., “Achieving k-anonymity privacy protection using generalization and suppression”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 2002, 10: 571- S88.
11. Li Z, Ye X. “Privacy protection on multiple sensitive attributes”.[Cl// Proceedings of the 9th international conference on information and communications security. Zhengzhou, China: Springer-Verlag; 2007: 141-152.
12. Zhong S, Yang Z, Chen T. “k-Anonymous data coliection”[J]. Information Sciences 2009, 179: 2948-2963.
13. L. Sweeney, “k-Anonymity: A model for protecting privacy”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5):557-570, 2002.
14. A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam,” l-Diversity: privacy beyond k-anonymity,” ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1), Article 3, 2007.
15. Wong R C, Li J Y, Fu A W, Wang K, ” (α,k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing”,Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '06, ACM, New York, USA, 2006. 754-759.106-115.
16. N. Li, T. Li and S. Venkatasubramanian.” t-Closeness: privacy beyond k-anonymity and l-diversity”, In IEEE ICDE 2007, pp. 106-115. IEEE, 2007.
17. P. Samarati. “Protecting respondents' identities in microdata release”, IEEE Transactions on Knowledge and Data Engineering, 13(6):1010-1027, 2001.
18. P. Samarati and L. Sweeney, “ Generalizing data to provide anonymity when disclosing information”, In Proc. of the 17th ACM SIGACTSIGMOD-SIGART Symposium on Principles of Database Systems (PODS 98), Seattle, WA, June 1998, p. 188.
19. Sara Hajian and Josep Domingo- Ferrer, “A methodology for Direct and Indirect Discrimination Prevention in Data Mining,” IEEE Trans. Knowledge and Data Eng., vol. 25, no. 7, pp. 1445-1459, July 2013.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

20. 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.
21. D.J. Newman, S. Hettich, C.L. Blake, and C.J. Merz, "UCI Repository of Machine Learning Databases," http://archive. ics.uci.edu/ml,1998.
22. R. Kohavi and B. Becker, "UCI Repository of Machine Learning Databases," http://archive.ics.uci.edu/ml/datasets/Adult, 1996.

Sreejith S., received B.Tech degree in Information Technology from M. S. University and M.E in Computer and Communication Engineering from Anna University. He is currently working as an Assistant Professor in Computer Science and Engineering Dept. at LBSITW, Thiruvananthapuram, India.

Sujitha S., received B.Tech degree in Computer Science and Engineering from University of Kerala in 2011. At present she is a PG Scholar at LBS Institute of Technology for Women, Thiruvananthapuram, India.