Discrimination Prevention with Privacy Preservation in Data Mining

Shilpa Gonbare[1], Satishkumar Varma[2], Manjusha Deshmukh[3]

[1,3] Department of Computer Engineering, PIIT, New Panvel, India
[2]Department of Information Technology, PIIT, New Panvel, India

***Abstract*— *Discrimination is a crucial issue while considering the legality and morality of data mining. People never like to be distinguished on basis of their caste, gender, nationality, etc. specifically while making decisions with consideration of this attributes, like offering them a job, property, loan, and so forth. Also when a data is released for analysis, there are chances that personal data may be misused for different reasons. Therefore, many companies and institutions are trying to bridge the gap between mining and sharing user data so that their services are improved and user privacy is ensured. We are trying to adapt an approach which will publish the discrimination free data such that probability of learning sensitive value of individual will be reduced. Concept of Slicing is used for preserving privacy. Slicing is better approach of privacy publishing in comparison with other approaches like generalization and bucketization, as it preserves correlation between attributes and it prevents attribute disclosure. The input to the system is discriminated dataset and output will be the sliced data that helps to preserve privacy.*

***Keywords*— Discimination, discrimination discovery, discrimination prevention, privacy, slicing**

I. INTRODUCTION

Discrimination is a very sensitive issue in data mining. Discrimination means to differentiate between. Discrimination leads to unbiased treatment of an individual just because they belong to a certain category. For example, one may be discriminated on basis of one's race, sexual category, nationality and so on (also called as discriminatory attributes) especially when making decisions like offering them a job, house, insurance, finance, and so on.

Discrimination is commonly classified into two types, direct or indirect. Rules or procedures that give indication about deprived groups, based on discriminatory items related to particular group are known as direct discriminatory rules. Rules that, do not mention discriminatory items, but may generate discriminatory results are known as indirect discriminatory rules.

Classification rules [7] in data mining uses historical data for producing auto decisions, like approval/rejection of loan, personal selection etc. But if the historical data is prejudiced towards a particular community then the output will be showing discriminatory behaviour. Discovery and elimination of such possible biases from training data is very much required. Discrimination prevention techniques are used to prevent such type of discrimination. To avoid such discrimination, discrimination techniques such as discrimination discovery of discrimination and prevention of discrimination have been introduced in data mining.

Also when a data is released for analysis, privacy preservation techniques are very much essential, as they reduce the chances of identifying sensitive information about individuals, which they do not want to disclose. For example, in medical data, sensitive information can be the disease of a particular patient he/she suffers from. Now users are becoming more and more concerned about their privacy; the way their personal data is used and getting shared. So sharing personal data ensuring user's privacy and preventing discrimination is very much important today.

II. LITERATURE SURVEY

In this section, we mention the earlier literature that uses the various discrimination prevention and privacy preserving data publishing techniques. For discrimination prevention, most of the researchers transformed discriminatory data in such a way that discrimination was avoided. Also there are many privacy preserving data publishing techniques like generalization, bucketization although slicing is more powerful than those.

Authors [1] addressed the problem of discrimination in data mining. They gave idea about how discrimination can be discovered by measuring the discrimination using a measure known as generalization of lift. They have introduced concept of α-protection which is nothing but a threshold used to make decision about whether the classification rules containing one or more discriminatory items is discriminatory or non-discriminatory. The results were tested on German credit data set. They have identified α-discriminatory rules which satisfy minimum support of 1%, 0.5% and 0.3% and α values altering from 1.4 to 2.0.

Authors [2] presents an approach consisting of messaging the data so that discrimination is removed with minimum possible changes. The class labels of the most probable victims (discriminated ones) and profiteers (favored ones) was changed by them. Naive Bayesian classifier ranker is used to determine the victims and profiteers. This ranker ranks the data objects in accordance with their probability of being in the target class. The training data is modified until the data becomes discrimination free. It has disadvantages like it works only for direct discrimination and applicable to only one discriminatory attribute.

The paper [3] proposed a method for indirect discrimination prevention which considers several discriminatory attributes. They first discovered whether there exists indirect discrimination. In the event of any indirect discrimination, the modification of dataset is done until discrimination is decreased to a certain threshold or is totally eliminated. The method consists of transforming the original data in such that each redlining rule gets converted to non-redlining rule with minimum loss of information. The transformed dataset is evaluated with measures like Discrimination Prevention Degree, Discrimination Protection Preservation. It has limits like, it works for only indirect discrimination and there is little information loss.

S. Hajian and J. Domingo-Ferrer [4] overcome the above limitations by proposing new techniques that are applicable for preventing direct and indirect discrimination separately or both at the same time. Dataset is transformed so that direct and/or indirect discriminatory decision rules gets converted to non-discriminatory classification rules. Success of the techniques in removal of direct and/or indirect discrimination from the original data set, and the impact of the procedure in terms of loss of information was measured , which shows that the techniques are not only effective but also preserves data quality.

Alphonsa Vedangi and V. Anandam [5] have presented a new technique called slicing for privacy-preserving publishing. Slicing approach consist of partitioning the data horizontally and vertically by preserving correlation between attributes. Slicing has many other features like it is applicable to high dimensional data, prevents attribute disclosure. The experiment was done on medical data.

III. PROBLEM STATEMENT

The idea is to create a discrimination free dataset and to publish data in such a way that individual's privacy is conserved. It starts with discovering the discrimination in dataset, removing it using discrimination prevention technique and then applying slicing technique for publishing data so that privacy is preserved. The input to the system is discriminated dataset and output will be sliced data that helps to preserve privacy.

IV. METHODOLOGY

The idea is to create a discrimination free dataset and to publish data in such a way that probability of identifying individuals sensitive value gets reduced i.e privacy of individual is preserved. It starts with discovering the discrimination in dataset, removing it using discrimination prevention technique and then applying slicing technique for publishing data so that privacy is preserved. The input to the system is discriminated dataset and output will be sliced data that helps to preserve privacy. Existing method discovers the discrimination and remove discrimination by data transformation method.

In the proposed approach as shown in Figure 1 we will publish discrimination free data in such a way that individual's privacy is conserved. So a technique called slicing is used for privacy preservation.

The approach consists of three phases:

1. Discrimination discovery
2. Discrimination Prevention
3. Privacy preservation using slicing

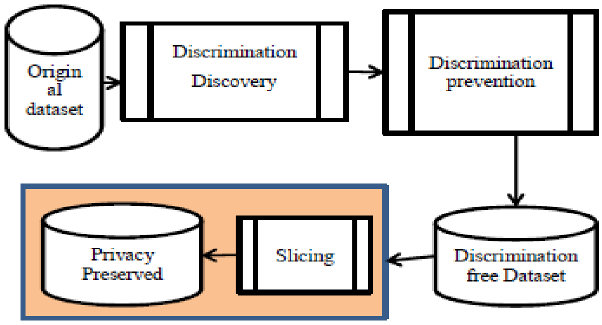


Fig.1: Proposed Approach

As shown in figure 1, first step is to discover the discriminatory situation in the dataset. In case if any discrimination is found, then it is removed using data transformation methods. Finally slicing technique is applied, which partitions the dataset horizontally and vertically by preserving correlation between attributes. This technique publishes data in such a way that probability of identifying individuals sensitive value gets reduce.

4.1. Discrimination Discovery

Let DI be the discriminatory itemset we have selected. A classification rule $X \rightarrow C$ is called potentially discriminatory (PD) when X is consisting of discriminatory item set.

A classification rule $X \rightarrow C$ is called non discriminatory when X does not contain discriminatory item set. Direct discrimination is measured using direct discrimination measure (elift) and discriminatory threshold (α). Indirect discrimination is measured using indirect discrimination measure (elb) and discriminatory threshold (α).

The Steps in discrimination discovery are given below:

1. This phase accepts discriminatory dataset as input. Discriminatory items are used to generate frequent classification rules.
2. These frequent classification rules are categorized into PD (potentially discriminated) and PND (potentially non-discriminated) rules.
3. Identification of Direct discrimination is done by determining α -discriminatory rules among the PD rules. These α -discriminatory rules are determined with the help of a direct discrimination measure and a discriminatory threshold (α).
4. Identification of indirect discrimination is done by recognizing redlining rules among the PND rules. These rules are determined with the help of an indirect discriminatory measure and a discriminatory threshold (α).

4.2. Discrimination Prevention

Ones discrimination is identified, it is removed using discrimination prevention technique. The original dataset is transformed so that direct and indirect discriminatory biases are removed with least effect on the data and on permissible decision rules. It converts each α -discriminatory rule into an α -protective rule and each redlining rule into non-redlining rule, based on the direct discriminatory measure and indirect discriminatory measure respectively.

4.3. Privacy Preservation using slicing

Slicing segments the dataset both vertically and horizontally. It consists of three phases: tuple partitioning, attribute partitioning and column generalization.

1. In tuple partitioning, tuples are group into buckets.
2. After tuple partitioning, the correlations between pairs of attributes is computed and then attributes are clustered based on their correlations so that correlations among attributes is preserved.
3. Finally column generalization is applied. Columns in each bucket are generalized into less specific but logically consistent values.

V. EXPERIMENTS

5.1. DATA SETS

Adult data set [8]: This data set consists of 48,842 records, train dataset is of 5000 records. The data set has 14 attributes (without class attribute). The prediction task associated with the Adult data set is to determine whether a person makes more than 50K$ a year based on census and demographic information about people. The data set contains both categorical and numerical attributes. Age is the discriminatory attribute in the dataset which is numerical, we converted it to categorical by partitioning its domain into two fixed intervals: Age <= 30 is considered as Young and Age > 30 is renamed as Old.

German credit data set [9]: This data set consists of 1,000 records and 20 attributes (without class attribute) of bank account holders. This is a very well-known real-life data set, comprising of both numerical and categorical attributes. The class attribute in the German Credit data set takes values representing good or bad classification of the bank account holders.

5.2. UTILITY MEASURES

We have evaluated discrimination prevention on the base of following metrics:

- Direct discrimination prevention degree (DDPD).

This measure calculates the percentage of α-discriminatory rules that are no longer α-discriminatory in the transformed data set.

- Direct discrimination protection preservation (DDPP).

This measure measures the percentage of the α-protective rules in the original data set and transformed data set

- Indirect discrimination prevention degree (IDPD).

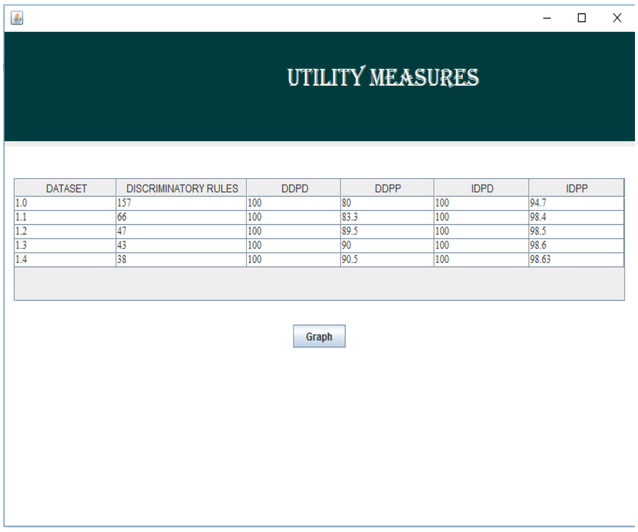
This measure measures the percentage of redlining rules that are no longer redlining in the transformed data set.

- Indirect discrimination protection preservation (IDPP).

This measure measures the percentage of nonredlining rules in the original data set and transformed data set.

5.3 RESULTS

Our main focus is to eliminate both direct and indirect discrimination simultaneously, so we have implemented the algorithm, and we have evaluated it in terms of utility measures with respect to different α (threshold) values.



DATASET	DISCRIMINATORY RULES	DDPD	DDPP	IDPD	IDPP
1.0	157	100	80	100	94.7
1.1	66	100	83.3	100	98.4
1.2	47	100	89.5	100	98.5
1.3	43	100	90	100	98.6
1.4	38	100	90.5	100	98.63

Fig.2: Result of Adult Dataset

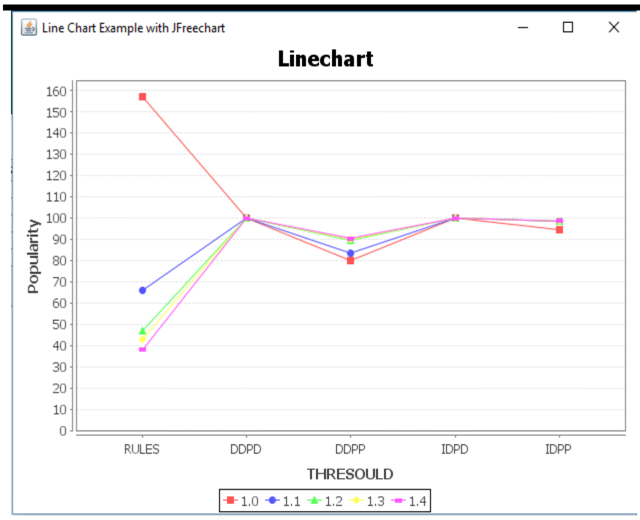


Fig.3: Graph Representing Result of Adult Dataset

In Adult dataset , by considering discriminatory attributes as Age, Race, Sex, we get result as shown in Figure 2 and Figure 3 for α varying from 1.0 to 1.4. We can see that number of discriminatory rules decreases as threshold (α) value increases. Also percentage of utility measures increases as threshold (α) value increases.

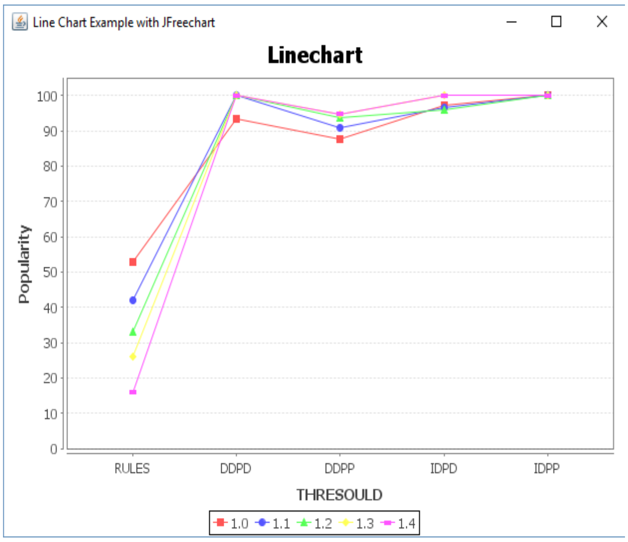
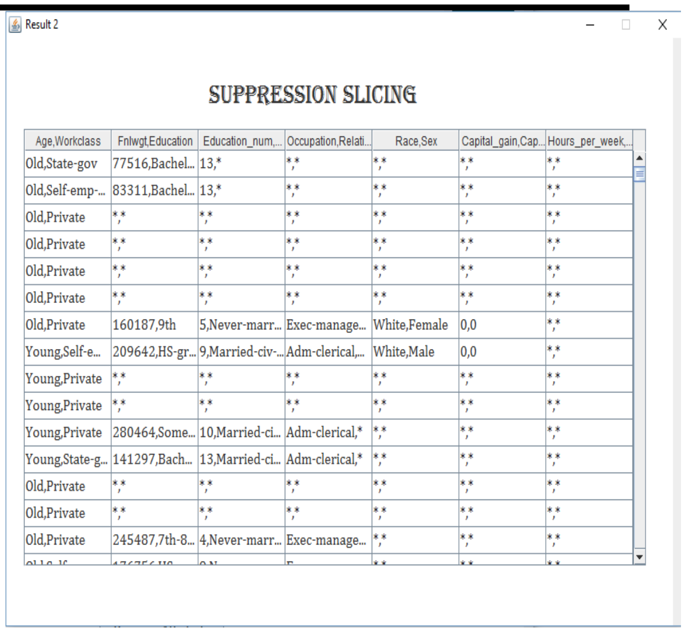


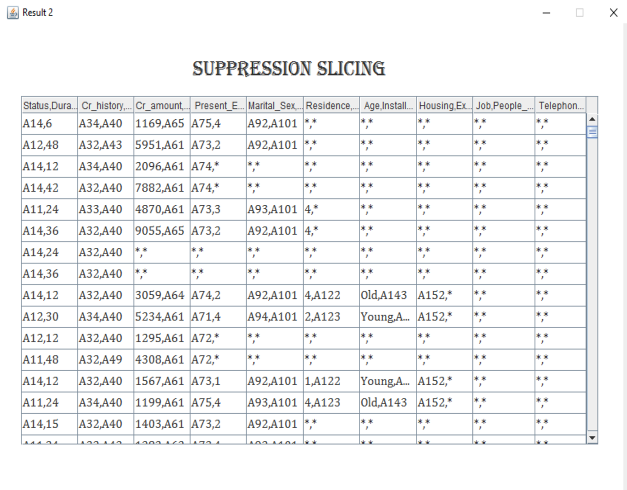
Fig.4: Graph Representing Result of German Credit Dataset

In German Credit dataset , by considering discriminatory attributes as Foreign_worker, Marital_sex, Status, we get result as shown in Figure 4 for α varying from 1.0 to 1.4. We can see that number of discriminatory rules decreases as threshold (α) value increases. Also percentage of utility measures increases as threshold (α) value increases.



Age,Workclass	Fnlwgt,Education	Education_num,...	Occupation,Relati...	Race,Sex	Capital_gain,Cap...	Hours_per_week,...
Old,State-gov	77516,Bachel...	13,*	*,*	*,*	*,*	*,*
Old,Self-emp-...	83311,Bachel...	13,*	*,*	*,*	*,*	*,*
Old,Private	*,*	*,*	*,*	*,*	*,*	*,*
Old,Private	*,*	*,*	*,*	*,*	*,*	*,*
Old,Private	*,*	*,*	*,*	*,*	*,*	*,*
Old,Private	*,*	*,*	*,*	*,*	*,*	*,*
Old,Private	160187,9th	5,Never-marr...	Exec-manage...	White,Female	0,0	*,*
Young,Self-e...	209642,HS-gr...	9,Married-civ-...	Adm-clerical,...	White,Male	0,0	*,*
Young,Private	*,*	*,*	*,*	*,*	*,*	*,*
Young,Private	*,*	*,*	*,*	*,*	*,*	*,*
Young,Private	280464,Some...	10,Married-ci...	Adm-clerical,*	*,*	*,*	*,*
Young,State-g...	141297,Bach...	13,Married-ci...	Adm-clerical,*	*,*	*,*	*,*
Old,Private	*,*	*,*	*,*	*,*	*,*	*,*
Old,Private	*,*	*,*	*,*	*,*	*,*	*,*
Old,Private	245487,7th-8...	4,Never-marr...	Exec-manage...	*,*	*,*	*,*

Fig.5: Slicing of discrimination free adult dataset



Status,Dura...	Cr_history,...	Cr_amount,...	Present_E...	Marital_Sex,...	Residence,...	Age,Install...	Housing,Ex...	Job,People_...	Telephon...
A14,6	A34,A40	1169,A65	A75,4	A92,A101	*,*	*,*	*,*	*,*	*,*
A12,48	A32,A43	5951,A61	A73,2	A92,A101	*,*	*,*	*,*	*,*	*,*
A14,12	A34,A40	2096,A61	A74,*	*,*	*,*	*,*	*,*	*,*	*,*
A14,42	A32,A40	7882,A61	A74,*	*,*	*,*	*,*	*,*	*,*	*,*
A11,24	A33,A40	4870,A61	A73,3	A93,A101	4,*	*,*	*,*	*,*	*,*
A14,36	A32,A40	9055,A65	A73,2	A92,A101	4,*	*,*	*,*	*,*	*,*
A14,24	A32,A40	*,*	*,*	*,*	*,*	*,*	*,*	*,*	*,*
A14,36	A32,A40	*,*	*,*	*,*	*,*	*,*	*,*	*,*	*,*
A14,12	A32,A40	3059,A64	A74,2	A92,A101	4,A122	Old,A143	A152,*	*,*	*,*
A12,30	A34,A40	5234,A61	A71,4	A94,A101	2,A123	Young,A...	A152,*	*,*	*,*
A12,12	A32,A40	1295,A61	A72,*	*,*	*,*	*,*	*,*	*,*	*,*
A11,48	A32,A49	4308,A61	A72,*	*,*	*,*	*,*	*,*	*,*	*,*
A14,12	A32,A40	1567,A61	A73,1	A92,A101	1,A122	Young,A...	A152,*	*,*	*,*
A11,24	A34,A40	1199,A61	A75,4	A93,A101	4,A123	Old,A143	A152,*	*,*	*,*
A14,15	A32,A40	1403,A61	A73,2	A92,A101	*,*	*,*	*,*	*,*	*,*

Fig.6: Slicing of discrimination free german credit dataset

Discrimination free dataset is sliced as shown in Figure 5 and Figure 6 so that privacy of the data is preserved.

VI. CONCLUSION

The proposed system focused to avoid two negative issues relates to data mining that is Discrimination and Privacy invasion. The purpose of the paper is to make the dataset free of discrimination as well as to preserve the privacy of data. To achieve this first discrimination is identified and then data is transformed so that all the biases the dataset are removed. Finally, discrimination free dataset is generated. Also privacy of the dataset is preserved by representing data so that chances of identifying individuals data gets reduced. The experiment results show that we are achieving the goal of removing discrimination and privacy preservation.

REFERENCES

[1] D. Pedreschi, S. Ruggieri, and F. Turini, "Discrimination-Aware Data Mining," Proc. 4th ACM International Conference Knowledge Discovery and Data Mining (KDD '08), pp. 560-568, 2008.
[2] F. Kamiran and T. Calders, "Classification without Discrimination," Proc. IEEE Second International Conference Computer, Control and Communication (IC4 '09), 2009.
[3] S. Hajian, J. Domingo-Ferrer, and A.Martı´nez-Balleste´, "Rule Protection for Indirect Discrimination Prevention in Data Mining" Proc. Eighth International Conference Modeling Decisions for Artificial Intelligence (MDAI '11), pp. 211-222, 2011.
[4] S.Hajian and Josep Domingo-Ferrer, "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining", IEEE Transactions On Knowledge And Data Engineering, Vol. 25,No. 7, July 2013.
[5] Alphonsa Vedangi, V. Anandam, " Data Slicing Technique to privacy preserving And Data publishing" International journal of research in engineering and technology, VOL. 2, Issue. 10, Oct-2013.
[6] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "ℓ-diversity: Privacy beyond k-anonymity". In International council for open and distance education, page 24, 2006.
[7] M.H. Dunham, "Data Mining Introductory and Advanced Topics", Pearson Education.
[8] R. Kohavi and B. Becker, "UCI Repository of Machine Learning Databases," http://archive.ics.uci.edu/ml/datasets/Adult, 1996.
[9] 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.