



7th International Conference on Communication, Computing and Virtualization 2016

Discrimination Prevention with Classification and Privacy Preservation in Data mining

Krishna Kumar Tripathi^a

Sant Gadge Baba Amravati University, Amravati, MS, India, 444601

Abstract

Mining of data is an important increasingly methodology for extracting and finding the meaningful hidden knowledge in huge archives of data. There are the various negative social perceptions related to mining of data, out of which many are potential discrimination and potential privacy invasion. The potential discrimination consists of unfairly treating and identifying people based on their existence and belonging to a particular group. Data mining and automated data collection techniques such as classification and association rule mining have provided way to taking decisions automatically, such as computation of insurance premium, loan granting or denial, credit card issue etc. If the provided data sets for training are biased in discriminatory (sensitive) attributes such as, race, gender, religion, etc., discriminatory decisions can be taken and may ensue. For avoiding this situations, antidiscrimination methodology like discrimination prevention and discovery have been considered in the data mining. There are mainly two types of discrimination, one is direct discrimination and second is indirect discrimination. Direct discrimination exists in the situations when decisions are taken on the basis of the sensitive attributes. Indirect discrimination exists in the situations when decisions are taken on the basis of the non-sensitive attributes that are strongly correlated with the biased sensitive attributes.

© 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).
Peer-review under responsibility of the Organizing Committee of ICCCV 2016

Keywords: *Discrimination; correlation; Antidiscrimination*

I. INTRODUCTION

In human society, discrimination is the detrimental treatment of the people on the basis of their membership or existence in a certain category or group. It involves restriction/prohibition to members of one group opportunities that exists to the other groups and categories. There are various types of antidiscrimination acts, that are created in order to restrict discrimination based on the number of attributes (e.g., race, gender, religion, disability, nationality, marital status, and age) in various settings (E.g. Training and employment, access to public services, insurance and credit, etc.). For instance, the European Union have implemented the principle of the equal treatment among men and women in the supply and access to services and goods [1] or in matters of occupation and employment in [2]. Even there are

the laws exists against discrimination, all of them are basically reactive and not proactive. Emerging technologies and research can add proactivity to laws by helping in the discrimination discovery techniques and prevention mechanisms. Information and services in the society allows an automatic and routine collection of huge amounts of data used for mining. These data are often used in order to train the classification/ association rules in view of taking the automated decisions in various process including, loan granting/rejection, Computation of insurance premium, personnel selection, credit card issue, etc. At the initial stages, automating decisions making may give the sense of Fairness: classification/ association rules do not leads themselves by the personal preferences. But, by looking closely, we can realizes that the classification/association rules are mainly learned by the system which is machine learning (e.g., loan rejection/ granting) from the provided training data. If the provided training data are biased in nature for or against a particular category/ group / community (e.g., foreigners), then in that case the technically learned model may show a discriminatory detrimental results/behavior.

In the other words, the devolved model may suggest that just being foreign is valid reason for denying of loan. Discovering and identifying such potential biases and removing them from the provided training data without damaging their decision making capability is therefore main requirement. We should prevent data mining modal / techniques from becoming itself as a source of discrimination, Data mining tasks resulting discriminatory models based on the biased data sets the part of making automated decision. In [5], it is presented that the data mining can be both a means for discovering discrimination and a source of discrimination.

1.1 Aim of study

Firstly, a review of literature on the aspects of discrimination prevention and privacy preservation is conducted to address the issue of detrimental treatment of the people on the basis of sensitive attributes, no major work with proper discrimination prevention is carried out in this area and generated the motivation to carryout the research in this area.

Secondly, the review of documents, reports, other literature & analysis of current systems suggested that existing systems do not have enhanced discrimination prevention & privacy preservation mechanisms

Main aim of this paper is to propose architecture to develop a method to make the original data both privacy protected and discrimination-free. [6]

1.2 Objective

The prime objective and requirement of the thesis work is to develop a system that provides discrimination prevention as well as privacy preservation with classification and clustering of the data .With the proposed work we are planning to achieve the following objectives:. [10]

- Improvement in the existing approach
- Obtaining better results for discrimination prevention and privacy preservation
- Creating a new method for the discrimination prevention
- Classification of the training data set
- Finding frequent association rules present in the training data set
- Clustering of the training data set
- Using randomization approach for differential privacy preservation
- Comparing the proposed method with the existing methods.

II LITERATURE SURVEY

Despite having the wide deployment of the information systems on data mining technology in the way of decision making, discriminations present in data mining did not receive much of attention until 2008. Hence, along with discrimination discovery in data mining, even more challenging issue is to restrict knowledge-based decision support systems from taking discriminatory decisions. Various approaches and techniques resembles on the discovery and measurement of the discrimination. The various other techniques deals with discrimination prevention. [12]

The discrimination discovery mechanism was first proposed by Pedreschi et al. [4], [13]. This methodology is based

on the mining of categorization rules of the training data (the inductive part) and analysis them (the deductive part) on the basis of root of quantitative actions of discrimination which does the formalization of authorized definitions of the discriminations. In the US Equal Pay Act [14] mentions that: “the rate of selection for any given ethnic sex and race which is less than four-fifths of the rate for the given set having the highest rate will normally be regarded as the evidence of conflicting effect”. In the existing discrimination discovery techniques consider every rule used is independently used for the measurement of discrimination without accepting other rules or having the relation among themselves. In this thesis we also consider relation between rules for discrimination discovery present in the training data, depending on the absence or presence of discriminatory attributes. In the discrimination prevention techniques, the other main antidiscrimination requirement in data mining is the introduction of the patterns which do not leads to the discriminatory decisions even when unique training data sets are biased.

2.1 Analysis of the problem

During the study and investigation of the above literature survey, some of the issues were explored and are summarized by using the below mentioned points:

- The relationship between privacy preservation and discrimination prevention in mining of data is not investigated. It remains untouched topic of research whether privacy preservation can help in the anti-discrimination or vice-versa.
- The technique focuses on the attempt to identify discrimination in the original training data only for one discriminatory item and also based on a single measuring technique
- They doesn't include any measure to identify and evaluate how much discrimination has been removed from the training data and the amount of information loss has been incurred.
- The synergies among rule hiding for discrimination removal and rule hiding in privacy preserving data mining is not evaluated.
- They focus either on the direct discrimination prevention or indirect discrimination or not on both.
- The techniques also doesn't show any methods to calculate the amount of discrimination that has been removed, and hence do not concentrate on the amount of information loss which is generated.

Hence the proposed thesis work in data mining introduces various data mining techniques which overcome the above limitations. And also propose new data transformation methods (rule generalization (RG) and rule protection) are depends on the evaluation for both direct and indirect discrimination and also can deal with various discriminatory items [8].

2.2 Privacy Aware Data Mining Proposal

Privacy is basically not just a service or goal like security, but it is the belief of the people's to reach a controllable and protected state, possibly even without having to actively monitor for it by themselves. Hence, privacy is described as “the rights of the people's or individuals to recognize for themselves what, when and how information about them is used for various goal and purpose”. The preservation and protection of responsive data is an essential area of research that has involved many researchers in field of information technology. In the discovery of the discovery, attempt at assuring privacy when sharing and mining personal data have led to introducing privacy preserving in data mining (PPDM) methodologies.

PPDM have become more popular as they allow sharing and publishing sensitive data for the secondary level analysis. Various PPDM models (measures) and techniques and have been suggested to tradeoffs the service of the resulting models/ data for protecting individual privacy against different kinds of privacy attacks. [10]

III.METHODOLOGY FOR FUNCTIONING

The proposed web-based platform discrimination prevention with classification and privacy preservation emphasizes on the collaborative effort in modeling automated discrimination prevention, classification, and clustering, mining association rule and privacy preservation.

The methodology for developing the solution to the above discussed problem is presented in this section. The step by step approach of discrimination prevention with classification and privacy preservation under the case study is presented.

1. Critical review of the literature indiscrimination prevention, classification, clustering, mining association rule and privacy preservation.
2. Study the existing discrimination prevention techniques & privacy preservation and to identify the problems associated with it.
3. Training data collection to analyze the discriminations associated with it.
4. Design a framework for web-based automated system for discrimination prevention with classification and privacy preservation.
5. Design a folder structure for training data and processed data storage.
6. Select appropriate web clients' server architecture (2-tier/3-tier) and a data base system (centralized/decentralized) for implementation of the prototype system.
7. Develop an algorithm for training dataset parsing training data proposing, rule generalization and classification and clustering of the data.
8. Implement the prototype system and test the developed framework.

Services offered by the proposed web-based platform discrimination prevention with classification and privacy preservation

1. Introduction of data discrimination and privacy preservation
2. Uploading and processing the dataset
3. Rule generalization (RG) and rule protection for nominal attributes
4. Dataset Preprocessing
5. Data Discretization for numeric attributes
6. Extracting the preprocessed data in excel sheet
7. Classification of the preprocessing data
8. Clustering of the preprocessing data
9. Mining frequent association rules from the preprocessing data
10. Differential privacy preservation using randomization
11. Dataset Post processing
12. Extracting the post processed data in excel sheet
13. Storing the preprocessed, discretized and post processed data in to the database
14. Reporting and analytics section
15. Experimental results
16. Performance measure
17. Authentication mechanisms for admin user
18. Change password and Profile view for the admin user

IV. DATA SET

German Credit dataset which can be obtained from <ftp://ics.uci.edu/pub/machine-learning-databases/statlog/> is used for the above proposed web based thesis work. German Credit data set totally contains 1000 records, 13 nominal and 7 numeric attributes, with credit as a class label that can be good or bad. This dataset is most frequently used in the literature anti-discrimination. We are using nominal attribute for proposed web based dissertation work. The

Predictive task associated with the German credit dataset is to determine whether a record/person is granted a credit (good) or denied a credit (bad). Personal Status = Female and Foreign Worker = Yes and no single are treated as discriminatory attributes. Attribute Job can be treated as a sensitive attribute. Personal status, Age, Foreign worker, Property magnitude, Own Telephone can be considered as QIs (Quasi-identifier attributes).

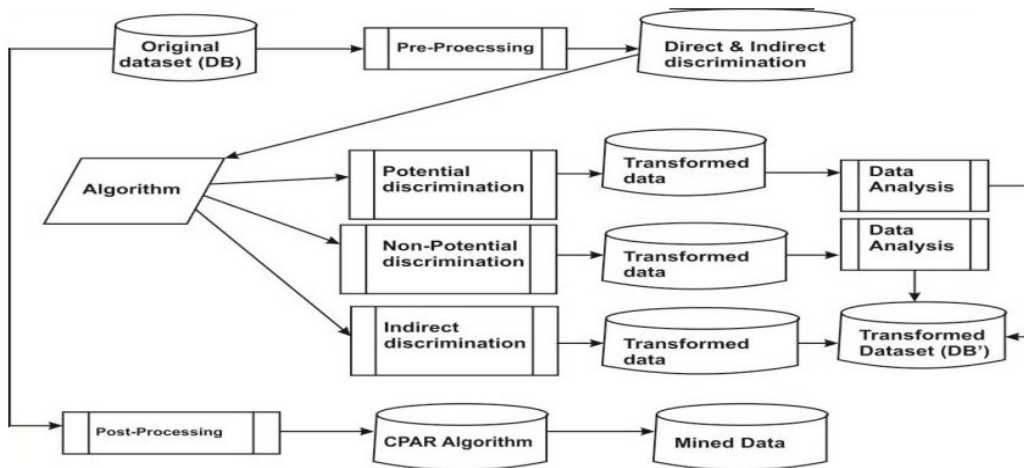


Fig 1 System Architecture

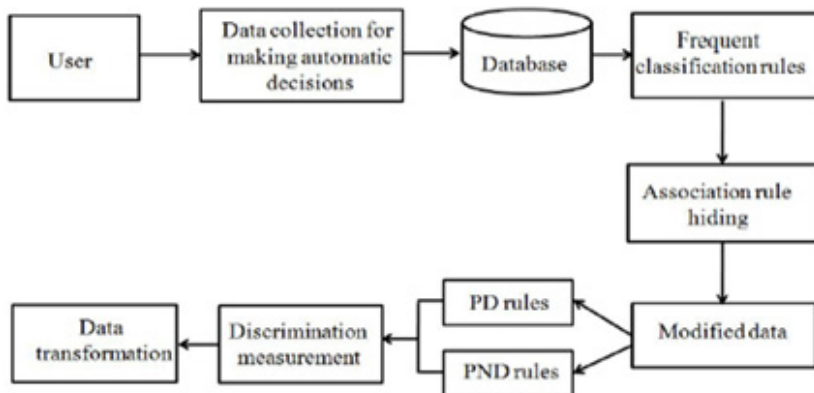


Fig 2 System Classification

Inputs to our proposed web based system are mentioned below-

Sensitive attribute (s) of the German credit data set: these are the attributes in the input dataset that contain sensitive details/information (e.g. salary).

Quasi-identifier attributes (QI) of the German credit data set: these are the set of attributes in the input dataset that can be used to re-identify/ assure an individual/ peoples (e.g. gender, age that can be used to re-identify a person’s salary).

Discriminatory attribute (DA) of the German credit data set: these are the set of attributes in the input dataset that are specified as discriminatory by nature/law (e.g. race, gender etc.)[6]

V. EXPERIMENT RESULT

The functioning and operation ability of each of the service is discussed in details in the following section.

1. Loading Data module will browse the data set and upload the data set to the framework, once the data set is uploaded it will fetch and display, Uploaded file name, Uploaded file size, Upload time, Total attributes, Total records

Along with that it will display the data set attribute information with following column

- Attribute name
- Description
- Possible values

German Credit Data Set
 The German Credit data set (available at [ftp.ics.uci.edu/pub/machine-learning-databases/statlog/](http://ics.uci.edu/pub/machine-learning-databases/statlog/)) contains observations on 30 variables for 1000 past applicants for credit. Each applicant was rated as good credit (700 cases) or bad credit (300 cases).
 Each applicant was rated as good credit (700 cases) or bad credit (300 cases).

Load the german data set :

credit_data_set.txt
 Data set uploaded successfully

File name : credit_data_set.txt
 File size : 151.0 KB
 Upload time : 01/10/2016 15:58:13
 Total Attributes : 21
 Total Records : 1000

German credit data set attribute information			
Index	Attribute name	Description	Possible values
1	Checking status	Status of existing checking account	<0, 0<=X<200, >=200, no checking account
2	Duration	Duration in month	Numeric
3	Credit history	Credit history	no credits taken/all credits paid back duly, all credits at this bank paid back duly, existing credits paid back duly bill now, delay in paying off in the past, critical account/other credits existing (not at this bank)
4	Purpose	Purpose	new car, used car, furniture/equipment, radio/tv, domestic appliance, repairs, education, vacation, retraining, business, other
5	Credit amount	Credit amount	Numeric
6	Savings status	Savings account/bonds	<100, 100<=X<500, 500<=X<1000, >=1000, unknown/no savings account
7	Employment	Present employment since	unemployed, <1, 1<=X<4, 4<=X<7, >=7
8	Installment commitment	Installment rate in percentage of disposable income	Numeric
9	Personal Status	Personal status and gender	male
10	Other debtors / guarantors	Other debtors / guarantors	none, co applicant, guarantor
11	Residence_since	Date beginning permanent residence	Numeric
12	Property magnitude	Property	real estate, building society savings agreement/life insurance, car, no known property
13	Age	Age in years	Numeric
14	Other payment/installment plans	Other installment plans	bank, stores, none

Fig 3 Loading data set

2. Rule generalization module takes one or a few facts about the data and making a broader, more universal statement for every records. It's the methodology of formulating general concepts by abstracting the common properties of instances form the training dataset. Rule generalization will apply on the Nominal Attribute of the dataset.
3. Data preprocessing module will transform the source/ training dataset in such a manner that the discriminatory biases contained in the original data are removed so that zero/no unfair decision rule can be mined/extracted from the transformed data in order to obtained error and discrimination free data
4. Discretization module will transform numeric attribute into range of data.

- Classification methodology is a data mining (machine learning) technique which used to predict group assignment's/membership for data instances of the training dataset. Assigning an object to a certain predefined class on the basis of its similarity to earlier examples of other objects it can be done by the reference to original training data or based on a model of that data. The main objective of the classification techniques is to analyze the input training data and to develop an accurate description or framework / model for each of the class using knowledge/features present in the data.

Rule generalization is taking one or a few facts and making a broader, more universal statement. It's the process of formulating general concepts by abstracting common Rule generalization will apply on the **Nominal Attribute of the dataset**

Attribute name and Generalized value (Original, Generalized)				
Checking status	<0	less	critical/other existing credit	Exists
	0<=X<200	medium	existing paid	Paid
	no checking	no checking	delayed previously	Delayed
	>=200	large	no credits/all paid	Clear
		all paid	Clear	
Purpose	radio/tv	Entertainment		
	education	Education		
	furniture/equipment	Equipment		
	new car	Vehicle		
	used car	Vehicle		
	business	Business		
	domestic appliance	Appliance		
	repairs	Clear		
	other	Other		
	retraining	Retraining		
		no known savings	No savings	
		<100	Low	
		500<=X<1000	Less	
		>=1000	High	
		100<=X<500	Medium	

Fig 4 Rule generalization

Attribute list	
Index	Attribute name
1	Checking status
2	Duration
3	Credit history
4	Purpose
5	Credit amount
6	Savings status
7	Employment
8	Installment commitment
9	Personal Status
10	Other debtors / guarantors

Name : checking_status Type : Nominal Distinct : 4 Missing : 0 (0 %)		
Index	label	Count
1	0 <= X < 200	269
2	< 0	274
3	> = 200	63
4	no checking	394

Fig5 Data preprocessing

This model is used to classify training data for which the class descriptions are not known.

For a frequent association rule referred as $X \rightarrow Y$, the support measure is denoted as $\text{sup}(X \rightarrow Y)$ which is the number of records/transactions where XUY observed divided by total number of records/transactions.

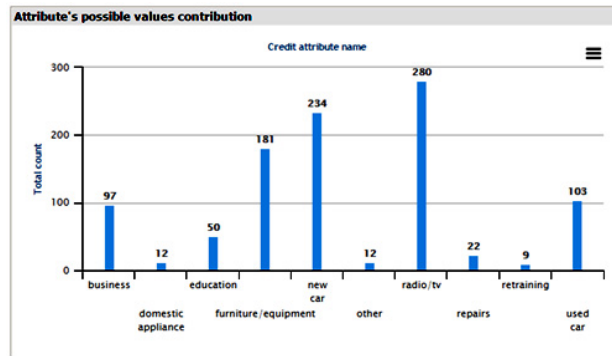


Fig 6 Data Attribute values

Attribute list		Classification of attribute : Purpose			
Index	Attribute name	Attribute label	Count	Support	Confidence
1	Checking status	business	97		
2	Duration	business --> good	63	0.06	0.65
3	Credit history	business --> bad	34	0.03	0.35
4	Purpose	domestic appliance	12		
5	Credit amount	domestic appliance --> good	8	0.01	0.67
6	Savings status	domestic appliance --> bad	4	0.0	0.33
7	Employment	education	50		
8	Installment commitment	education --> good	28	0.03	0.56
9	Personal Status	education --> bad	22	0.02	0.44
10	Other debtors / guarantors	furniture/equipment	181		
11	Residence_since	furniture/equipment --> good	123	0.12	0.68
12	Property magnitude	furniture/equipment --> bad	58	0.06	0.32
13	Age	new car	234		
14	Other payment/installment plans	new car --> good	145	0.14	0.62
15	Housing	new car --> bad	89	0.09	0.38
16	Existing_credits	other	12		
		other --> good	7	0.01	0.58
		other --> bad	5	0.0	0.42

Fig 7 Training data classification

The confidence measure is the number of records/transactions where XUY observed divided by the number of records/transactions where X appears.

$$\text{Confidence} = \text{Occurrence (Y)} / \text{Occurrence (X)}$$

$$\text{Support} = \text{Occurrence} / \text{Total Support}$$

The confidence measure of a rule $X \rightarrow Y$ is denoted as $\text{conf}(X \rightarrow Y) = \text{sup}(X \cup Y) / (\text{sup}(X))$. The support measure of

a rule $X \rightarrow Y$ is defined as $\text{sup}(X \rightarrow Y) = \text{sup}(XUY) / N$

N is the number of records in the transaction database,

$\text{Sup}(XUY)$ is the number of records containing X and Y ,

$\text{Sup}(X)$ is the number of records containing X

$\text{Sup}(Y)$ is the number of records containing Y .

VI. Algorithm

Input: Credit data set file

Output: (transformed data set)

1. Find item sets from DB
2. for each sensitive item h in H
3. If H is null then EXIT
4. Select a rule / record r from Association dataset file
5. Compute support and confidence of rule r
6. If $\text{conf} < \text{minconf}$ and $\text{supp} < \text{minsupp}$
7. Find $T_i = \{ t \text{ in DB } \mid t \text{ does not support and partially supports } r$
8. Change the value of sensitive item h
9. Compute support and confidence from rule r
10. Until (Association is empty)
11. } //end of if
12. Else
13. Compute Support and confidence of r
14. Update DB with new transaction t
15. Calculate confidence and support for class=good and class = bad
16. Plot a graph with x axis as 'attributes' and y axis as 'confidence values
17. Plot a graph with x axis as 'attributes' and y axis as 'support values

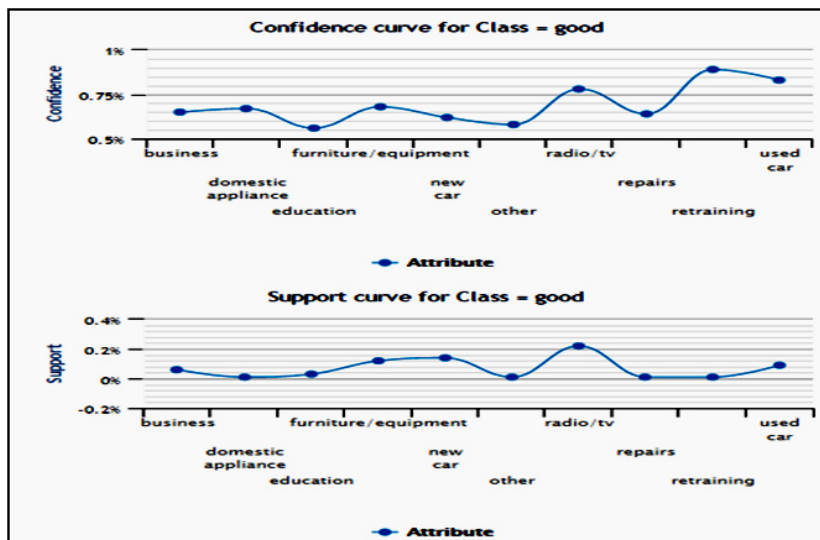


Fig 8 Confidence and Support

VI. CONCLUSION

The above solution proposed is the web-based framework for discrimination prevention with classification and privacy preservation. Here registered user can upload the training dataset, perform rule generalization (RG) and rule protection, perform data discretization, data preprocessing and classification

VI. FUTURE WORK

In the next paper, will elaborate how the proposed web based framework can be used for performing clustering, mining frequent association's rules, and privacy preservation and discrimination prevention along with the performance measure and experimental results.

REFERENCES

1. S. Hajian and J. Domingo, "A Methodology for Direct and Indirect Discrimination prevention in data mining." IEEE transaction on knowledge and data engineering, VOL. 25, NO. 7, pp. 1445-1459, JULY 2013.
2. S.R.Rathnapriya, J. Jebathangam and Dr. A. Muthukumaravel, "A methodology for direct and indirect Discrimination prevention in data mining" IJISSET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 6, August 2014.
3. M. Wakchaure and Prof. Dr. S. Sane "Discrimination Prevention by Different Measures in Direct Rule Protection Algorithm" www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 5, Issue 9, September 2015).
4. Dino Pedreschi, Salvatore Ruggieri and Franco Turini, "Discrimination-aware Data Mining" Dipartimento di Informatica, Università di Pisa L.go B. Pontecorvo 3, 56127 Pisa, Italy.
5. European Commission, "EU Directive 2006/54/EC on Anti-Discrimination," <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:204:0023:0036:en:PDF>, 2006.
6. Asmita Kashid, Vrushali Kulkarni and Ruhi Patankar, "Discrimination Prevention using Privacy Preserving Techniques" International Journal of Computer Applications (0975 – 8887) Volume 120 – No.1, June 2015.
7. Neenu Mary Kuruvila and V. Vennila, "International Journal of Advanced Research in Computer Science and Software Engineering" Volume 3, Issue 12, December 2013.
8. Anjali P and Renji S, "Discrimination Prevention in Data mining with Privacy Preservation" International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013).
9. S. Hajian and J. Domingo, "Generalization-based Privacy Preservation and Discrimination Prevention in Data Publishing and Mining".
10. Kamal D. Kotapalle and Shyam Gupta, "Discrimination Prevention and Privacy Preservation in Data Mining" www.ijird.com July, 2014 Vol 3 Issue 7 INTERNATIONAL JOURNAL.
11. Priya Meshram and Prof. Sonali Bodkhe, "Review on Privacy Preservation Method by Applying Discrimination Rules in Data Mining" International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 12, December 2014 ISSN: 2278 – 7798 .
12. Harshali M. Saindane and Mrs. V. L. Kolhe, "Survey of Approaches for Discrimination Prevention in Data Mining" International Journal of Computer Science and Information Technologies, Vol. 5 (6), 2014.
13. David H. Author, "The Economics of Discrimination —Theory" November 24, 2003.
14. Bart Goethals and Mohammed J. Saki, "Advances in Frequent Item set Mining Implementations: Report on FIMI'03" FIMI Repository: <http://fimi.cs.helsinki.fi/>