ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Direct and Indirect Discrimination Prevention Approach using Association Rule Hiding

Harshali M. Saindane[1], V. L. Kolhe[2]

P. G. Student, Dept. of Computer Engineering, D Y Patil College of Engineering, Akurdi, SPPU, Pune, India[1]

Assistant Professor, Dept. of Computer Engineering, D Y Patil College of Engineering, Akurdi, SPPU, Pune, India[2]

ABSTRACT: Discrimination prevention and Privacy preservation are amongst the important challenges in data mining. Association rule hiding is a privacy preservation technique that can be used to hide sensitive items for ensuring safety from privacy threats. The proposed work deals with using this privacy preservation technique to improve the discrimination prevention system. Discrimination is said to occur when people are treated unfairly on the basis of their sensitive attributes like gender, religion, race etc. It is of two types namely direct and indirect. Direct discrimination consists of biased decision rules based on sensitive attributes like religion, race, community etc. Indirect discrimination can occur when decision rules are based on non-sensitive attributes which are closely related to sensitive ones. Data mining methods like classification rule mining are used for making automated decisions by decision support systems for personnel selection, loan sanctioning etc. Discriminatory decisions may occur if the training data sets are biased with respect to the sensitive attributes. Discrimination discovery and prevention have been introduced in data mining to generate legal decision rules. The proposed system consists of using association rule hiding concept by embedding it in the existing discrimination prevention system based on rule protection and generalization. This system can be used to prevent discrimination more effectively. It can be helpful to preserve original dataset quality.

KEYWORDS: Data Mining, Discrimination prevention, Direct and indirect discrimination, privacy preservation, discrimination measures, Association Rule Hiding

I. INTRODUCTION

Discrimination is the illegal treatment to individuals based on their sensitive features like religion, race etc in sociology [1]. It involves denying opportunities to people of one group that are available to other group of people. Although antidiscrimination laws exist, discrimination is still a prevalent and challenging issue in data mining. Now a days, automated decision support systems are used to reduce the workload of staff of an organization for applications like personnel selection, loan sanctioning etc [2]. These systems are based on classification rule mining technique of data mining. Since classification model learns from the training data, if the training dataset is inherently biased for a certain group, discrimination is likely to occur. Discrimination is of two types namely direct and indirect. Direct discrimination leads to biased decision rules obtained from sensitive discriminatory attributes like gender, religion etc. Indirect discrimination leads to biased decision rules based on non-sensitive attributes closely related to the sensitive ones. Along with discrimination discovery, ensuring that knowledge-based decision support systems do not infer discriminatory decisions is a more challenging issue. The challenge increases when there is a need to prevent not only direct discrimination but also indirect discrimination or both.

Although there are some approaches for each of the above mentioned method, discrimination prevention has become the latest research avenue. Discrimination prevention based on pre-processing approach seems the most flexible, since it does not require changing the standard data mining algorithms, unlike the in-processing approach, and it allows data publishing unlike the post-processing approach. The proposed system is based on pre-processing approach.

The rest of the paper is organized as follows: Section II depicts information about related work including different approaches used for discrimination prevention. Section III contains implementation details that includes system architecture, Association rule hiding algorithm and mathematical model for the system. Section IV describes the dataset

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

used for experimental purpose and evaluates the results . Section V contains the conclusion of research work and the future work.

II. RELATED WORK

Beyond discrimination discovery, preventing knowledge based decision support systems from making discriminatory decisions is a more challenging issue [1]. Despite of the large usage of information systems based on data mining in decision making, discrimination in data mining was not explored until 2008.

Dino Predeshi, S. Ruggieri F. Turini [3] are the first researchers to address the problem of discrimination in data mining models. This approach is based on identifying the discriminatory rules that are present in a dataset, and the specific subset of the data where they exist, rather than on learning a classifier with independency constraints for future predictions.

D Predeshi and S. Ruggieri [11] have also developed a systematic framework for measuring discrimination, analysing historical decision records stored out of a socially- sensitive decision task, e.g., credit approval. This work is an extension of their earlier work.

F. Kamiran and T. Calders [4] have introduced a Classification with No Discrimination model (CND). This method is based on massaging the dataset by making the least intrusive modifications leading to a dataset that is bias free. A non discriminating classifier is then learned on this modified dataset. This method classifies the future data with minimum discrimination and high accuracy. They have extended this approach by introducing a preferential sampling scheme [5] to make the data discrimination free. This scheme changes the distribution of different data objects for a given data to make it discrimination free.

Toon Calders and S. Verwer [12] have modified the Naive Bayes classifier in order to perform classification that is restricted to be independent with respect to a sensitive attribute. Three approaches for making the Naive Bayes classifier discrimination-free are 1) modifying the probability of the decision being positive, 2) training one model for every sensitive attribute value and balancing them, and 3) adding a latent variable to the Bayesian model that represents the unbiased label and optimizing the model parameters for likelihood using expectation maximization.

S. Ruggieri, D. Pedreschi, and F. Turini [6] have developed the DCUBE system based on classification rule extraction and analysis, using an Oracle database. DCUBE tool helps in guiding the users about the legal issues about discrimination hidden in data, and through several legally grounded analysis to deal with discriminatory situations.

F. Kamiran and T. Calders [7] have presented the construction of a decision tree classifier without discrimination. This is a different approach of addressing the discrimination aware classification problem. In this approach, the non-discriminatory constraint is pushed deeply into a decision tree learner by changing its splitting criterion and pruning strategy by using a novel leaf relabeling approach.

Sara Hajian and J. Domingo-Ferrer [8] have introduced anti-discrimination in the context of cyber security. They have developed a new discrimination prevention method based on data transformation that can consider several discriminatory attributes and their combinations. They have introduced some measures for evaluating this method in terms of its success in discrimination prevention and its impact on data quality. The limitation of this approach is that it deals only with direct discrimination.

Sara Hajian, Domingo-Ferrer and A. Martinez-Balleste [9] have introduced a new pre-processing method for indirect discrimination prevention based on data transformation that can consider several discriminatory attributes and their combinations. They have used some measures for evaluating their method in terms of its success in discrimination prevention and its impact on data quality. It is the first work that have addressed a discrimination prevention method for indirect discrimination.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

S. Hajian and J. Domingo-Ferrer [1] have addressed the problem of direct and indirect discrimination prevention in data mining. They have developed rule protection and generalization methods applicable for direct or indirect discrimination prevention individually or simultaneously. This method uses various methods to clean training data sets and outsourced data sets in such a way that direct and/or indirect discriminatory decision rules are converted to legal classification rules.

Association Rule hiding is the process of hiding strong association rules and creating sanitized database from the original database in order to prevent unauthorised party to generating frequent sensitive patterns[14]. Association rule hiding algorithms can be divided into three different approaches. They are heuristic approaches, border-revision approaches and exact approaches[14].

Kasthuri S. and Meyyappan T. et al. have proposed a heuristic based approach for association rule hiding[15]. This approach selects all the association rules containing sensitive items either in the left or in the right from the set of all association rules generated from a dataset. These rules are represented in representative rules (RR) format with sensitive item on the left hand side or right hand side of the rules. A rule is selected from the set of RRs which contains sensitive item. Then a transaction is selected which completely supports RR, which means that it contains all the sensitive items in the RR. The proposed approach hides the sensitive item by modifying the database without changing the support of the sensitive item.

Thus based on the above literature survey, methods based on rule protection and generalization are effective in discrimination prevention. The proposed system based on association rule hiding can help improving the efficiency of existing system thereby maintaining original data quality.

III. IMPLEMENTATION DETAILS

The main objective of the proposed system is to improve the efficiency of discrimination prevention while maintaining original data with minimum information loss.

A. *System Architecture*

Fig.1. shows the proposed system architecture for Direct and Indirect Discrimination Removal using Association Rule Hiding method. The system takes training dataset as input and generates transformed dataset which is free from discrimination.

1) ***Finding frequent classification rules***:- The first step consists of finding frequent classification rules for the training dataset using Apriori algorithm.
2) ***Association rule hiding***:- The frequent rules from first step are used by association rule hiding's heuristic approach. Using it, if the given condition is satisfied, the values of sensitive items are changed. The confidence and support measure values are calculated up to two iterations. This approach generates sanitized data that is used to generate potentially discriminatory and potentially non-discriminatory rules.
3) ***PD and PND rules***:- The sanitized data is used to generate PD and PND rules based on predetermined discriminatory items in dataset.
    - PD Rules:- $X \rightarrow C$ classification rule is potentially discriminatory(PD) with X = A, B when $A \subseteq$ DIs a nonempty discriminatory item set and B is a non discriminatory item set.
    - PND Rules:- $X \rightarrow C$ is potentially non discriminatory (PND) classification rule with X = D, B is a non discriminatory item set.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

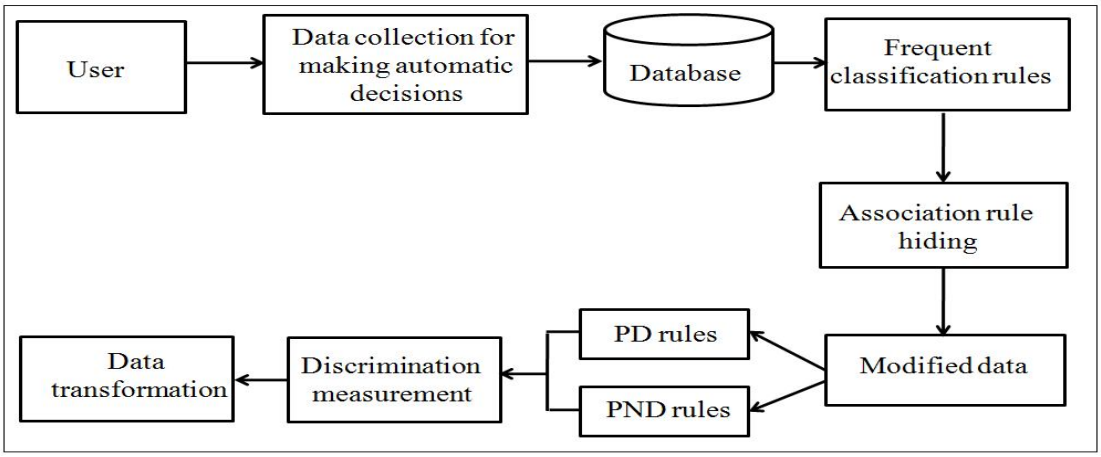

Fig.1. Proposed System Architecture

4) ***Discrimination measurement*:-** The first step deals with measuring discrimination in the training dataset. Direct and indirect discrimination discovery includes identifying α-discriminatory rules and redlining rules. Second, direct discrimination is measured by identifying α-discriminatory rules among the PD rules using a direct discrimination measure (elift) and a discriminatory threshold (α). Indirect discrimination is measured by identifying redlining rules among the PND rules combined with background knowledge, using an indirect discriminatory measure (elb), and a discriminatory threshold (α).

5) ***Data transformation*:-** Transform the original data in such a way to remove direct and/or indirect discriminatory biases that have minimum impact on the data and on legitimate decision rules, so that no unfair decision rule can be extracted from the transformed data. The transformed dataset will be more effective in terms of discrimination removal. It maintains original data quality thereby reducing information loss.

B. *Algorithm:*

Aim of the proposed algorithm is to improve the efficiency of existing discrimination prevention system thereby maintaining original data loss. The proposed algorithm consists of following pseudo code:-

Input : DB, H: set of sensitive items, minsupp, minconf
Output : DB' (transformed data set)
1. Find item sets from DB
2.For each sensitive item h ∈ H
3. If H is null then EXIT
4. Select a rule r from Association
5. Compute support and confidence of rule r
6. if conf < minconf and supp < minsupp
7. Find Ti = { t in DB | t 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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

15. End
16. Output: DB' = DB

C. *Mathematical Model*

- Input to the system
  I = Training dataset. The training dataset is used to generate frequent classification rules. Frequent rules are generated using Apriori algorithm.
- Finding frequent rules with sensitive items and transforming them
  T_d = Transformed dataset that consist of modifying frequent rules to find discrimination more effectively.
- Potentially discriminatory (PD) and non-discriminatory rules (PND):- The T_d dataset is used to generate PD and PND rules.
  R = {PD, PND}

  i) A classification rule A, B → C is PD rule if A is a discriminatory item while B is non-discriminatory.

  ii) A classification rule with D, B → C is PND rule if X= {D, B} is a non-discriminatory item set.

- Discrimination Measurement

$\delta = \{elift, elb\}$

a) elift (extended lift) is direct discrimination measure. Let A, B → C be a classification rule such that conf (B → C) > 0.

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

PD classification rule c = A, B → C is α-protective, if elift(c) < α else c is α-discriminatory.

b) elb is indirect discrimination measure used to identify redlining rules that are rules indirectly inferred from non-discriminatory items. Let r : D, B → C be a PND classification rule, and let γ = conf(rb1:D, B → C) and δ = conf(B → C) > 0.

Let A be a discriminatory item set, and let β_1, β_2 such that conf(rb1 : A, B → D) ≥ β_1, conf(rb1 : D, B → A) ≥ β_2 > 0.

$$f(x) = \frac{\beta_1}{\beta_2}(\beta_2 + x - 1) \quad (2)$$

$$elb(x, y) = \begin{cases} \frac{f(x)}{y}, & \text{if } f(x) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

For α ≥ 0, if elb(γ, δ) ≥ α, the PD classification rule of form, r': A, B → C is α-discriminatory.

- Direct Indirect Discrimination Prevention Method
  M = {DRP, IRP}

i. DRP = Direct Rule Protection method :- The DRP method2 alters the class item ¬C to C for some records that satisfy the condition, ¬A, B → C.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

ii. IRP = Indirect Rule Protection method :- IRP method2 alters the class item $\neg$ C to C for some records that satisfy the condition, $\neg$ A, B, $\neg$ D $\rightarrow$ C.

- Performance Measures

  P = {DDPD, DDPP, IDPD, IDPP}

a) DDPD = Direct discrimination prevention degree measure It quantifies the percentage of α-discriminatory rules that are no longer α-discriminatory in the transformed data set.

$$\text{DDPD} = \frac{|MR| - |\text{MR}'|}{|MR|} \quad (4)$$

b) DDPP = Direct discrimination protection preservation measure It quantifies the percentage of α-protective rules in original dataset that remain α- protective in the transformed data set.

$$\text{DDPP} = \frac{|PR| \cap |\text{PR}'|}{|PR|} \quad (5)$$

c) IDPD = Indirect discrimination prevention degree measure quantifies the percentage of redlining rules that are no longer redlining in the transformed data set.

d) IDPP = Indirect discrimination protection preservation quantifies the percentage of non-redlining rules in the original data set that remain non-redlining in the transformed data set.

- Output

  O = {D'}

  O is the set of outputs from the system. D' is the set of transformed dataset which is free from direct and indirect discrimination.

IV. RESULTS AND DISCUSSION

A. *DataSet*

Standard dataset is used for experimental purpose. Adult data set is used, which is also known as Census Income. It comprises of 48,842 records. It 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 dataset is available at the UCI repository. It is popular dataset used for data mining tasks like classification. The data set has been downloaded from [16].

B. *Result Set*

The proposed system have been implemented for direct and indirect discrimination prevention using various methods like DRP (Method1), DRP (Method2), DRP combined with RG method for preventing direct discrimination and DIDP method for simultaneous direct and indirect discrimination prevention. Then we compared both approaches, Discrimination Prevention without Association Rule Hiding and Discrimination Prevention with Association Rule Hiding. The experimental setup includes 1500 records from Adult dataset. The minimum support value is 5% and minimum confidence value is 10%. The value of discriminatory threshold i.e. α is set to 1.0. Table I shows the results obtained for discrimination prevention without using Association Rule Hiding method. It gives the number of frequent rules, number of direct discriminatory, indirect discriminatory and redlining rules.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Table I: Results for Discrimination Prevention without using Association Rule Hiding

Method	*No. of frequent rules*	*No. of direct discriminatory rules*	*No. of indirect discriminatory rules*	*No. of redlining rules*
DRP (Method1)	1003	62	not applicable	not applicable
DRP (Method2)	1003	63	not applicable	not applicable
DRP (Method2)+ RG	1003	78	not applicable	not applicable
DIDP	1003	78	66	34

Table II shows the results obtained for the proposed system which is discrimination prevention using Association Rule Hiding method. It can be observed that the proposed system shows improved results for discrimination prevention. It helps in removing discrimination more effectively than existing system.

Table II: Results for Discrimination Prevention using Association Rule Hiding

Method	*No. of frequent rules*	*No. of direct discriminatory rules*	*No. of indirect discriminatory rules*	*No. of redlining rules*
DRP (Method1)	1080	70	not applicable	not applicable
DRP (Method2)	1080	71	not applicable	not applicable
DRP (Method2)+ RG	1080	86	not applicable	not applicable
DIDP	1080	86	70	38

We conclude that proposed system generate more efficient number of direct and indirect discrimination rules, after embedding with Association Rule Hiding method. The important point is that, by applying the proposed method, we get good results for both direct and indirect discrimination prevention at the same time. In addition, the values of MC and GC demonstrate that the proposed solution incurs low information loss.

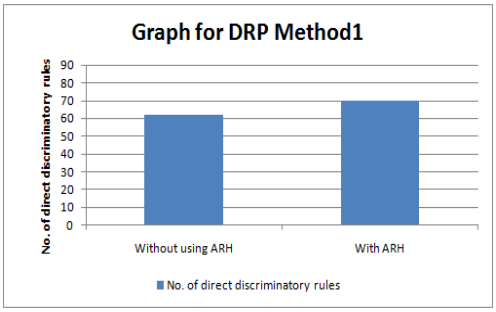

Fig.2. Graph for DRP Method1

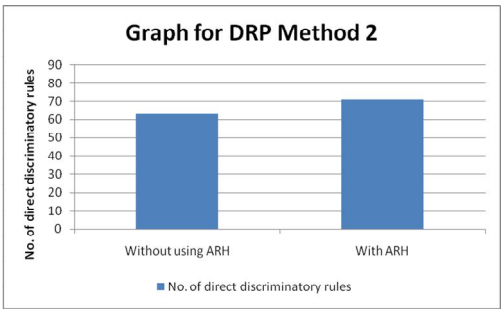

Fig.3. Graph for DRP Method2

Fig.2 and fig.3 show graph plotted for DRP Method1 and DRP Method2 respectively according to values of Table I and II. It can be seen that no. of direct discriminatory rules in proposed system are more than that of existing system.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

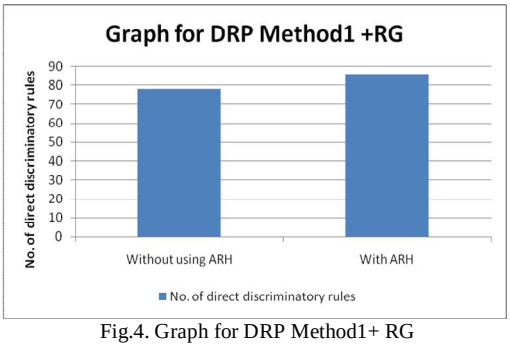

Fig.4. Graph for DRP Method1+ RG

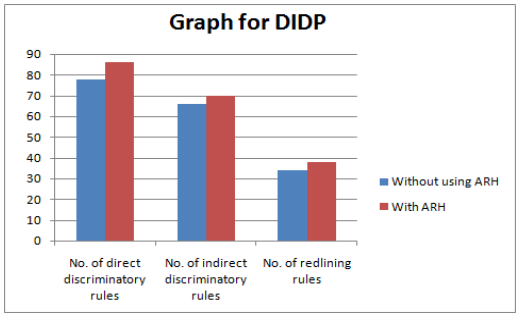

Fig.5. Graph for DIDP

Fig.4 and fig.5 show graph plotted for DRP Method1 with RG and DIDP method resp. according to values of Table I and II. DRP Method1+RG for proposed system shows more no. of direct discriminatory rules than existing system. Also DIDP method for proposed system is more effective in removing simultaneous direct and indirect discrimination than existing system.

V. CONCLUSION AND FUTURE WORK

In this paper, we presented an unified approach for simultaneous direct and indirect discrimination prevention based on Association rule hiding method. The main feature that distinguishes our approach from others is that privacy preservation data mining algorithm used for hiding sensitive data is used for discrimination prevention purpose. The objective is to improve the existing discrimination prevention system in terms of discovering direct and indirect discriminatory rules.

It is observed that unified approach for discrimination prevention is more effective than other methods, since it helps in removing direct and/or indirect discrimination at the same time and maintains original data quality.

For future work we intend to explore measures of discrimination different from the used measures (elift and elb). Also it would be interesting to find the discrimination system's relation with the current privacy models like differential privacy.

REFERENCES

1. S. Hajian and J. Domingo-Ferrer, "A methodology for direct and indirect discrimination prevention in data mining", IEEE Transactions on Knowledge and Data Engineering, vol. 25, pp. 1445-1459, 2013.
2. Hajian, Sara, and Josep Domingo-Ferrer. "Direct and indirect discrimination prevention methods", in Discrimination and Privacy in the Information Society, pp. 241-254, 2013.
3. D. Pedreshi, S. Ruggieri F. Turini, "Discrimination-aware data mining", in Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 560-568, 2008.
4. F. Kamiran and T. Calders, "Classifying without discriminating", IEEE 2nd International Conference on Computer, Control and Communication, pp. 1-6, 2009.
5. F. Kamiran and T. Calders, "Classification with no discrimination by preferential sampling", in Proc. 19th Machine Learning Conf., 2010.
6. S. Ruggieri, D. Pedreschi, and F. Turini, "Dcube: Discrimination discovery in databases", in Proceedings of the ACM SIGMOD International Conference on Management of data, pp. 1127-1130, 2010.
7. F. Kamiran, T. Calders, and M. Pechenizkiy, "Discrimination aware decision tree learning", IEEE 10th International Conference on Data Mining, pp. 869-874, 2010.
8. S. Hajian, J. Domingo-Ferrer, and A. Martinez-Balleste, "Discrimination prevention in data mining for intrusion and crime detection" , IEEE Symposium on Computational Intelligence in Cyber Security (CICS), pp. 47-54, 2011.
9. S. Hajian, J. Domingo-Ferrer, and A. Martinez-Balleste, "Rule protection for indirect discrimination prevention in data mining", in Modeling Decision for Artificial Intelligence, pp. 211-222, 2011.
10. D. Pedreschi, S. Ruggieri, and F. Turini, "The discovery of discrimination", in Discrimination and Privacy in the Information Society, pp. 91-108, 2013.
11. D. Pedreschi, S. Ruggieri, and F. Turini, "Measuring discrimination in socially-sensitive decision records", in SDM, pp. 581-592, 2009.
12. 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.
13. S. Ruggieri, D. Pedreschi, and F. Turini, "Data mining for discrimination discovery", ACM Transactions on Knowledge Discovery from data, vol.4, no.2, 9, 2010.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

14. Garg, Vikram, Anju Singh, and Divakar Singh. "A Survey of Association Rule Hiding Algorithms", IEEE 4th International Conference on Communication Systems and Network Technologies (CSNT), pp.404-407, 2014.
15. Kasthuri, S., and T. Meyyappan. "Hiding Sensitive Association Rule Using Heuristic Approach", International Journal of Data Mining & Knowledge Management Process, 2013.
16. R. Kohavi and B. Becker, "UCI Repository of Machine Learning Databases", http://archive.ics.uci.edu/ml/datasets/Adult, 1996.

BIOGRAPHY

Harshali M. Saindane received the BE degree in Computer Engineering from K.K.W.I.E.E.R, Nashik which is under Savitribai Phule Pune University, in 2008, and currently pursuing Master of Engineering degree in Computer Engineering from D. Y. Patil College of Engineering, Akurdi, Pune. Her research interests includes Data Mining and Soft Computing.

V. L. Kolhe is an Assistant Professor in the Computer Engineering Department, D. Y. Patil College of Engineering, Akurdi, Pune. She received the BE degree in Computer Technology from Nagpur University, and ME degree in Computer Engineering from Savitribai Phule Pune University.