ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

A Technology for Direct and Indirect Discrimination Prevention in Sensitive Data

A. Jagadeeswara Rao
Associate Professor & HOD,
Department of CSE,
Pace Institute Of Technology and Sciences,
Near Valluramma Temple.

M. Bhargavi
M.Tech Student,
Department of CSE,
Pace Institute Of Technology and Sciences,
Near Valluramma Temple.

S Phani Kumar
Assistant Professor,
Department of CSE,
Pace Institute Of Technology and Sciences,
Near Valluramma Temple.

Abstract:

Facts mining is an increasingly important technology for getting from useful knowledge put out of the way in greatly sized collections of knowledge for computers. There are however not social power of being conscious of about facts mining among which possible & unused quality right not to be public attack and take by force and possible & unused quality judgment.

The latter is chiefly of wrongly giving attention to people on the base of their being the property of to a special group made automatic knowledge for computers getting together and facts mining techniques such as order rule mining have covered the way to making made automatic decisions like loan giving agreement words saying not true insurance more than normal value computation and so on.

If the training facts puts are had a tendency in a certain direction in what in connection with discriminatory sensitive properties like sex race religion and so on discriminatory decisions may come after for this reason antidiscrimination techniques including judgment discovery and putting a stop to have been introduced in facts mining judgment can be either straight to or roundabout straight to judgment comes to mind when decisions are made based on sensitive properties roundabout judgment comes to mind when decisions are made based on no sensitive properties which are strongly connected with had a tendency in a certain direction sensitive ones.

In this paper we do judgment putting a stop to in facts mining and make offer new techniques able to be used for straight to or roundabout judgment putting a stop to one at a time or both at the same time.

We have a discussion how to clean training facts makes ready and outsourced knowledge for computers set in such a way that straight to and or roundabout discriminatory decision rules are converted to right nondiscriminatory order rules.

We also make an offer new metrics to value the use of the made an offer moves near and we make a comparison these moves near the testing values put examples on view that the made an offer techniques are working well at removing straight to and or round about judgment has a tendency in a certain direction in the first form knowledge for computers put while keeping safe facts quality.

1 Introduction:

In sociology judgment is the damaging process of a person based on their number of persons in a society in a certain group or sort. It gets into saying no to members of one group chances that are ready (to be used) to other groups. There is a list of anti discrimination acts which are laws designed to put a stop to judgment on the base of a number of properties e.g., do quickly religion sex ones's nation condition of being unable of married position (in society) and existence-stage in different gold frames e.g., use and training way in to public services credit and insurance and so on for example the European Union implements the sense of equal process between men and women in the way in to and supply of goods and services in or in matters of use and rule by strangers in although there are some laws against be acting differently to different persons all of them are reactive not before-the-fact technology can make an addition proactivity to (making) laws by sending in (writing) judgment discovery and putting a stop to techniques.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

Services in the information society let for automatic and regularly order group of greatly sized amounts of facts. Those facts are often used to train connection order rules in view of making made automatic decisions like loan giving agreement words saying not true insurance more than normal value computation personnel selection and so on. At first view automating decisions may give a sense of degree of shade order rules do not guide themselves by personal desires. However at a closer look one gets clear about that order rules are actually learned by the system e.g., loan giving agreement from the training facts.

If the training facts are inherently had a tendency in a certain direction for or against one town e.g., persons of another country the learned design to be copied may play or amusement a discriminatory bad feeling behavior. In other words the system may use reasoning that just being out-of-country is a right reason for loan words saying not true making discovery of such possible & unused quality has a tendency in a certain direction and taking away them from the training facts without causing damage their decision-making use is therefore highly desirable one must put a stop to knowledge for computers mining from becoming itself a starting point of judgment needing payment to knowledge for computers mining tasks producing discriminatory models from had a tendency in a certain direction facts puts as part of the made automatic decision making in it is put examples on view that facts mining can be both a starting point of judgment and a means for making discovery of judgment.

Judgment can be either straight to or roundabout also telephoned ordered straight to judgment is chiefly of rules or procedures that clearly, with detail say the name of not old enough in law to act or unhelpful groups based on sensitive discriminatory properties related to group number of persons in a society roundabout judgment is chiefly of rules or procedures that while not clearly, with detail saying the name of discriminatory properties purposely or without purpose could produce discriminatory decisions. Redlining by money business institutions saying no to grant mortgages or insurance in of a town areas they take into account as becoming less in value is an archetypal example of roundabout judgment although certainly not the only one with a small, little bad language of language for the purpose of compactness in this paper roundabout judgment will also be has relation to as redlining and rules causing roundabout judgment will be named redlining rules round about judgment could come about because of the able to use of some back knowledge rules for example that a certain quick general rule of behavior of a group is like to a becoming less in value area or an area with mostly black population. The back knowledge might be able to be got to from publicly ready (to be used) knowledge for computers e.g., numbering of persons knowledge for computers or might be got from the first form facts put itself because of the existence of nondiscriminatory properties that are highly connected with the sensitive ones in the first form knowledge for computers make ready.

2. Related Work:

Despite the wide placing of information systems based on facts mining technology in decision making the question under discussion of anti discrimination in facts mining did not get much attention until 2008 12. Some proposals are adjustment to events to the discovery and measure of be acting differently to different persons. Others amount with the putting a stop to of judgment.

The discovery of discriminatory decisions was first made an offer by Pedreschi et Al. The move near is based on mining order rules the way of discovery from examples part and reasoning on them the reasoning from facts part on the base of (able to be) measured measures of judgment that give fixed form to lawful clear outlines of judgment for example the us equal undergo punishment Act 18 states that a selection rate for any race sex or of divisions of man group which is less than four fifths of the rate for the group with the highest rate will generally be looked upon as facts supporting of going against force of meeting blow this move near has been stretched to go round statistical sense, value of the got from designs of judgment in and to reason about answering yes acting and unequal approval.

Moreover it has been instrumented as an Oracle-based person used by another in. Current judgment discovery methods take into account each rule one at a time for measuring be acting differently to different persons without giving thought to as other rules or the relation between them.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

However, in this paper we also take into account the relation between rules for judgment discovery, based on the existence or nonexistence of discriminatory properties.

• **Preprocesing.** Putting a stop to, the other major antidiscrimination purpose in facts mining, is chiefly of getting designs that do not lead to discriminatory decisions even if the first form training facts puts are had a tendency in a certain direction. Three moves near are able to be formed in mind: Preprocessing make great change the starting point knowledge for computers in such a way that the discriminatory has a tendency in a certain direction had within in the first form knowledge for computers are removed so that no hard decision rule can be mined from the greatly changed facts and send in name for any of the quality example facts mining algorithms. The preprocessing moves near of facts great change and hierarchy-based generality can be adjusted from the right not to be public process of making safe literature. Along this line, act a controlled distortion of the training facts from which a classifier is learned by making immeasurably intrusive adjustments leading to a not having a tendency in any direction facts put. The preprocessing move near is useful for applications in which a facts put should be made public and/or in which facts mining needs to be did also by outside parties (and not just by the facts owner).

• **In-processing.** Change the facts mining algorithms in such a way that the coming out models does not have within hard decision rules. For example, a that possibly taking place in addition move near to cleaning the judgment from the first form knowledge for computers put is made an offer in whereby the nondiscriminatory force to limit is fixed into a decision tree learner by changing its making into two rule for testing and pruning secret design through a new leaf relabeling move near. However, it is clearly and readily seen that in processing judgment putting a stop to methods must have belief in on new special-purpose facts mining algorithms; quality example knowledge for computers mining algorithms cannot be used.

• **Postprocessing.** Modify the coming out facts mining copies made to scale, instead of cleaning the first form knowledge for computers put or changing the facts mining algorithms. For example, in, a confidence-altering move near is made an offer for order rules worked out by the CPAR algorithm.

The after-processing move near does not let the facts put to be made public: only the made an adjustment facts mining models can be made public (knowledge putting into print), for this reason facts mining can be did by the facts owner only.

One might have in mind that of a straightforward preprocessing move near made up of just removing the discriminatory properties from the facts put. Although this would get answer to the straight to judgment hard question, it would cause much information loss and in general it would not get answer to roundabout judgment. As stated in there may be other properties (e.g., quickly) that are highly connected with the sensitive ones (e.g., group of the same blood) and let work out discriminatory rules. For this reason, there are two important questions looking upon judgment putting a stop to: one sporting offer is to take into account both straight to and roundabout judgment instead of only straight to judgment; the other sporting offer is to discover a good tradeoff between judgment be taken away and the quality of the coming out training facts puts and facts mining copies made to scale.

Although some methods have already been made an offer for each of the named beforehand moves near (preprocessing, in-processing, after-processing), be acting differently to different persons putting a stop to puts a stop to a largely unobserved research road. In this paper, we get, come together at one point on judgment putting a stop to based on preprocessing, because the preprocessing move near seems the most flexible one: it does not have need of changing the quality example facts mining algorithms, unlike the in processing move near, and it lets facts putting into print (more like than just knowledge putting into print), unlike the after-processing move near.

2.1 Contribution and Plan of This Paper:

Judgment putting a stop to methods based on preprocessing made public so far, present some limiting conditions, which we next mark:

• They attempt to discover judgment in the first form knowledge for computers only for one discriminatory one thing on a list and based on a single measure. This move near cannot be responsible for that the greatly changed knowledge for computers put is really judgment free,

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

because it is within one's knowledge that discriminatory behaviors can often be put out of the way behind several discriminatory things on a list, and even behind groups of them.

• They only take into account straight to judgment.

• They do not join any measure to value how much judgment has been removed and how much information loss has been caused.

In this paper, we make an offer preprocessing methods which overcome the above limiting conditions. Our new facts great change methods (i.e., rule system of care for trade and rule generality (RG)) are based on measures for both straight to and roundabout judgment and can business agreement with several discriminatory things on a list. In addition, we give use measures. For this reason, our move near to judgment putting a stop to is wider than in earlier work.

In our earlier work, we introduced the first idea of using rule system of care for trade and rule generality for straight to judgment putting a stop to, but we gave no based on experience outcomes. In, we introduced the use of rule system of care for trade in a different way for roundabout judgment putting a stop to and we gave some preliminary testing outcomes. In this paper, we present a joined move near to straight to and roundabout judgment putting a stop to, with finalized algorithms and all possible facts great change methods based on rule system of care for trade and/or rule generality that could be sent in name for straight to or round about judgment putting a stop to.

We specify the different features of each way. Since methods in our earlier papers, could only amount with either straight to or round about judgment, the methods described in this paper are new. As part of this hard work, we have undergone growth metrics that specify which records should be changed, how many records should be changed, and how those records should be changed during facts great change. In addition, we make an offer new company that does a public work measures to value the different made an offer be acting differently to different persons putting a stop to methods in terms of knowledge for computers quality and judgment be taken away for both straight to and roundabout judgment.

Based on the made an offer measures, we present much based on experience results for two noted facts puts and make a comparison the different possible methods for straight to or roundabout judgment putting a stop to discover out which methods could be more good in terms of low information loss and high be acting differently to different persons be taken away.

3 PROPOSED SYSTEM:

We propose new utility measures to evaluate the different proposed discrimination prevention methods in terms of data quality and discrimination removal for both direct and indirect discrimination. Based on the proposed measures, we present extensive experimental results for two well known data sets and compare the different possible methods for direct or indirect discrimination prevention to find out which methods could be more successful in terms of low information loss and high discrimination removal. The approach is based on mining classification rules (the inductive part) and reasoning on them (the deductive part) on the basis of quantitative measures of discrimination that formalize legal definitions of discrimination.

3.1 MODULE:
Number of Modules

After careful analysis the system has been identified to have the following modules:

1.Direct Discrimination Prevention Module.

2.Indirect Discrimination Prevention Module.

3.Rule Protection in Data Mining Module.

4.Rule Generalization in Data Mining Module.

1. Direct Discrimination Prevention Module:

Direct discrimination occurs when decisions are made based on sensitive attributes. It consists of rules or procedures that explicitly mention minority or disadvantaged groups based on sensitive discriminatory attributes related to group membership. To prevent direct discrimination is based on the fact that the data set of decision rules would be free of direct discrimination

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

if it only contained PD rules that are protective or are instances of at least one nonredlining PND rule. In this we apply direct rule protection and direct rule generalization.

2. Indirect Discrimination Prevention Module Module:

Indirect discrimination occurs when decisions are made based on nonsensitive attributes which are strongly correlated with biased sensitive ones. It consists of rules or procedures that, while not explicitly mentioning discriminatory attributes, intentionally or unintentionally could generate discriminatory decisions. To prevent indirect discrimination is based on the fact that the data set of decision rules would be free of indirect discrimination if it contained no redlining rules. To achieve this, a suitable data transformation with minimum information loss should be applied in such a way that redlining rules are converted to nonredlining rules. To overcome this we apply indirect rule protection and indirect rule generalization.

Direct and Indirect Discrimination Prevention Algorithms

Algorithm details: Our proposed data transformation method for simultaneous direct and indirect discrimination prevention. The algorithm starts with redlining rules. From each redlining rule (r : X C), more than one indirect _- discriminatory rule (r1: A,B C) might be generated because of two reasons: 1) existence of different ways to group the items in X into a context item set B and a nondiscriminatory item set D correlated to some discriminatory item set A; and 2) existence of more than one item in DIs. Hence, as shown in Algorithm (Step 5), given a redlining rule r, proper data transformation should be conducted for all indirect α- discriminatory rules r1 : (A DIs) (B X) C ensuing from r.

3.Rule Protection in Data Mining Module:

The data transformation is based on direct rule protection and indirect rule protection. classification rules do not guide themselves by personal preferences. However, at a closer look, one realizes that classification rules are actually learned by the system (e.g., loan granting) from the training data.

If the training data are inherently biased for or against a particular community (e.g., foreigners), the learned model may show a discriminatory prejudiced behavior. In other words, the system may infer that just being foreign is a legitimate reason for loan denial.

4. Rule Generalization In Data Mining Module:

The data transformation is based on direct rule generalization and indirect rule generalization. In rule generalization, we consider the relation between rules instead of discrimination measures. Assume that a complainant claims discrimination against foreign workers among applicants for a job position. In other words, foreign workers are rejected because of their low experience, not just because they are foreign. The general rule rejecting low-experienced applicants is a legitimate one, because experience can be considered a genuine/legitimate requirement for some jobs.

Algorithm. DIRECT AND INDIRECT DISCRIMINATION PREVENTION

1: Inputs: $\mathcal{DB}$, $\mathcal{FR}$, $\mathcal{RR}$, $\mathcal{MR}$, α, DI_s
2: Output: $\mathcal{DB}'$ (transformed data set)
3: **for** each $r : X \rightarrow C \in \mathcal{RR}$, where $D, B \subseteq X$ **do**
4: $\gamma = conf(r)$
5: **for** each $r' : (A \subseteq DI_s), (B \subseteq X) \rightarrow C \in \mathcal{RR}$ **do**
6: $\beta_2 = conf(r_{b2} : X \rightarrow A)$
7: $\Delta_1 = supp(r_{b2} : X \rightarrow A)$
8: $\delta = conf(B \rightarrow C)$
9: $\Delta_2 = supp(B \rightarrow A)$
10: $\beta_1 = \frac{\Delta_1}{\Delta_2}$ $//conf(r_{b1} : A, B \rightarrow D)$
11: Find $\mathcal{DB}_c$: all records in $\mathcal{DB}$ that completely support $\neg A, B, \neg D \rightarrow \neg C$
12: Steps 6-9 Algorithm 1
13: **if** $r' \in \mathcal{MR}$ **then**
14: **while** $(\delta \leq \frac{\beta_1(\beta_2+\gamma-1)}{\beta_2 \cdot \alpha})$ and $(\delta \leq \frac{conf(r')}{\alpha})$ **do**
15: Select first record db_c in $\mathcal{DB}_c$
16: Modify the class item of db_c from $\neg C$ to C in $\mathcal{DB}$
17: Recompute $\delta = conf(B \rightarrow C)$
18: **end while**
19: **else**
20: **while** $\delta \leq \frac{\beta_1(\beta_2+\gamma-1)}{\beta_2 \cdot \alpha}$ **do**
21: Steps 15-17 Algorithm 4
22: **end while**
23: **end if**
24: **end for**
25: **end for**
26: **for** each $r' : (A, B \rightarrow C) \in \mathcal{MR} \setminus \mathcal{RR}$ **do**
27: $\delta = conf(B \rightarrow C)$
28: Find $\mathcal{DB}_c$: all records in $\mathcal{DB}$ that completely support $\neg A, B \rightarrow \neg C$

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

29: Step 12
30: **while** $(\delta \leq \frac{conf(r')}{\alpha})$ **do**
31: Steps 15-17 Algorithm 4
32: **end while**
33: **end for**
34: Output: $\mathcal{DB}' = \mathcal{DB}$

5 CONCLUSIONS AND FUTURE WORK:

In company with right not to be public, judgement is a very important question under discussion when giving thought to as the lawful and right behavior aspects of facts mining? It is more than clearly and readily seen that most people do not need to be saw as different because of their sex, religion, ones's nation, age, and so on, especially when those properties are used for making decisions about them like giving them a mixed bag of goods, loan, insurance, and so on.

The purpose of this paper was to undergo growth a new preprocessing judgement putting a stop to methodology including different facts great change methods that can put a stop to straight to judgement, round about judgement or both of them at the same time.

To get to this end, the first step is to measure be acting differently to different persons and make out groups and groups of individuals that have been directly and/or in a roundabout way saw as different in the decision-making processes; the second step is to make great change facts in the right way to remove all those discriminatory has a tendency in a certain direction at last, judgement- free knowledge for computers models can be produced from the greatly changed facts put without seriously damaging knowledge for computers quality.

The testing results stated put examples on view that the made offer techniques are quite good in both goals of removing judgement and keeping safe facts quality. The power being conscious of judgement, just like the power being conscious of right not to be public, strongly depends on the lawful and art and learning agreements between groups of a society. Although we argued that judgement measures based on elift and elb are reasonable, as future work we make up one's mind to have a look for measures of judgement different from the ones thought out as in this paper.

This will have need of us to further work-room the lawful literature on judgement in several countries and, if importantly different judgement clear outlines and/or measures were to be discovered, new facts great change methods would need to be designed.

Last but not least, we need to have a look for the relation between judgement putting a stop to and right not to be public process of making safe in facts mining. It would be greatly interesting to discover synergies between rule keeping secret for privacy-preserving facts mining and rule putting out of the way for judgement is taken away. Just as we were able to make clear to that roundabout judgement be taken away can help straight to judgement be taken away, it remains to be seen whether right not to be public system of care for trade can help antidiscrimination or vice versa. The connection with current right not to be public copies made to scale, like be changing for different conditions right not to be public, is also a making secret designs research road.

REFERENCES:

[1] 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.

[2] 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.

[3] European Commission, "EU Directive 2004/113/EC on Anti- Discrimination," http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:373:0037:0043:EN:PDF, 2004.

[4] 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.

[5] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, "Discrimination Prevention in Data Mining for Intrusion and Crime Detection," Proc. IEEE Symp. Computational Intelligence in Cyber Security (CICS '11), pp. 47-54, 2011.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

[6] 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.

[7] F. Kamiran and T. Calders, "Classification without Discrimination," Proc. IEEE Second Int'l Conf. Computer, Control and Comm. (IC4 '09), 2009.

[8] F. Kamiran and T. Calders, "Classification with no Discrimination by Preferential Sampling," Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.

[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] R. Kohavi and B. Becker, "UCI Repository of Machine Learning Databases," http://archive.ics.uci.edu/ml/datasets/Adult, 1996.