



# **Review on Discrimination Detection and Prevention in Data Mining**

Pravinkumar S. Chavan<sup>1</sup>, Ram P. Tathe<sup>2</sup>, Akshay S. Thakare<sup>3</sup>.

UG Student, Dept. of CSE, Anuradha Engineering College, Chikhli, Maharashtra, India<sup>1</sup>

UG Student, Dept. of CSE, Anuradha Engineering College, Chikhli, Maharashtra, India<sup>2</sup>

UG Student, Dept. of CSE, Anuradha Engineering College, Chikhli, Maharashtra, India<sup>3</sup>

**ABSTRACT:** Data mining is the emerging technology which is mostly used in knowledge discovery and extraction of information from the available raw data. Due to the fact of quality classification rules and mining models the process of automated decision making has become very efficient. The process of automated decision making leads to the presence of discrimination in the results of mining models. Discrimination is the prejudicial treatment given to an individual by virtue of its association with certain group or categories. There are many factors on which an individual can be treated rationally such as belonging to some sensitive attributes as color, nationality, gender etc this is direct discrimination. There is another type which do not cause due to sensitive attributes but it indirectly depends upon sensitive attributes. Discrimination in data mining is also recognized as the potential problem of concern since automated decision making is employed everywhere now. Flaws in the trained historical data or tending towards or against certain group leads to the formation of discriminatory rules and finally the production of biased mining model.

**KEYWORDS:** Data Mining, Data transformation, Discrimination.

## **I. INTRODUCTION**

Data mining is the recent area of interest for researchers where it has the major emphasis on automated classification of data and decision making. Wherever automated decision making is considered then the system or model is always prone to discriminatory results. Discrimination is the unfair or distinguished treatment of people on the basis of the association with a particular group or membership with certain category in other words restricting one group of people and giving priority to another group is the discrimination [1].

Discrimination in data mining results due to the biased historical data which is trained towards or against certain group. The rules inferred from the biased trained data will ultimately results into biased classification rules and finally into a mining model which is discriminatory containing flaws and unfair decision making capabilities. Nowadays automated decision making is used everywhere in various fields such as computation of premium for insurance, job recruitment, loan approval/denial and so on. Hence it becomes evident to consider discrimination as the matter of concern in automated decision making and need to be eliminated considering all the related factors such as the loss of quality data and retaining the integrity of information. This paper discusses the basic outline of the discrimination prevention approaches and the review on the researches that took place in the same context. We will also discuss the taxonomy along with the basic techniques of discrimination prevention in data mining.

## **II. MOTIVATION**

The fact that prevention of discrimination on the basis of sensitive attributes as race, color, gender and nationality etc. is important aspect of automated decision making system. The basic solve over this problem can be a thought of data sanitization i.e. simply removing which are discrimination causing and leads to the biased classification rules and decision making. The disadvantage of data sanitization is that it may lead to loss of some useful quality information i.e. historic data and so data sanitization is not the solution we are looking for. Another feasibility is that one can transform the data so that it won't lead biased classification rules and mining model with flaws. A discrimination prevention technique consists of patterns that do not lead to discriminatory decisions even if the original



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

data set is biased[7]. Data transformation is considered to be a good solution for preventing discrimination and hence it is the optimal solution to develop a bias free mining model. DRP algorithm is an example for data transformation algorithm which provides fairly good results.

### III. LITERATURE REVIEW

Data mining is the area where researches are going at the great extent but still there are not much efforts reported in the field of discrimination detection and prevention. There are several reports which depicts the detection and measures to consider the level of discrimination but some of the proposals are there which quotes about the methodology for discrimination prevention.

The presence of discrimination [2] in classification rules and the mining models was discovered by D. Pedreschi. S. Ruggieri introduced a new method for measuring the extent of discrimination present in data mining model. F. Turini, D. Pedreschi and S. Ruggieri constructed a tool known as the DCUBE [8] which was the oracle based tool implemented to detect and measure the discrimination present in the data.

Sara Hajian and Josep Doming-Ferrer detected the flaws present in the existing mining models. Some of the limitations of existing models were:

- A) Discrimination was detected using a single attribute but there might be some cases that data is discriminated on the basis of more than one attribute. So considering only one attribute for discrimination prevention and detection was not optimal.
- B) In existing methodology only direct discrimination was considered indirect discrimination detection and prevention was nowhere the topic for consideration.
- C) No measure was there for the purpose of loss of quality information and data transformation due to the elimination of discriminatory attributes.

Since data sanitization was not optimal as mentioned earlier and hence Sara Hajian and Joseph Domingo Ferrer opted the method of data transformation i.e. rule generalization[1] for the purpose of preventing all three types of discrimination namely direct, indirect and the combination of both discrimination which was not present in earlier mining models.

F. Camiran, T. Calders and M. Pechenizkiy created a new method which was discrimination aware [4]. In their approach they trained the model with potentially biased data so that the model is aware of discriminatory data and will enhance the probability of making future decision fair.

For the construction of decision tree they proposed two new techniques to make it decision aware those techniques were:

- A) Construction of dependency aware decision tree :

At the time of constructing the decision tree the level of dependency is also evaluated along with the accuracy due to splitting criterion.

- B) Leaf relabeling :

Here only labels of the leaf are changed keeping the original data unaffected, hence no data is sacrificed and accuracy is preserved along with the minimum level of dependency.

T. Calders and S. Verwer introduced an approach using three Naïve Bayes classifier approach to determine the classification rules to design discrimination free mining models [5]. Three approaches were discussed in order to make Naïve Bayes classifier discrimination free they are:

- a) Modifying the probability of the decision being positive.
- b) Separate model for every sensitive attributes and balancing them.
- c) Adding the latent variable to Bayesian model.

It was not feasible to design a separate model for each sensitive attribute there might be large number of sensitive attributes in the database.

F. Kamiran and T. Calders introduced a new concept of classification with no discrimination(CND)[10].The method was to massaging the data with minimum possible transformation to the dataset to eliminate the discrimination, instead of relabeling the data objects or leaves in decision tree construction they provided us with the solution of sampling the data.

The author does not provided the model for discrimination prevention but it is accepted ethically the presence of small fraction of discrimination in the dataset. The major disadvantage of massaging the data is that it is highly



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

intrusive and hence preferential sampling (PS) [9] of the data was introduced by F. Kamiran and T. Calders. The preferential sampling method changes the distribution of data objects to make it discrimination free.

## IV. METHODS FOR DISCRIMINATION PREVENTION

There are several approaches for discrimination prevention but the method applied for the elimination of biased rules and knowledge discovery wholly depends on the type of discrimination it implies that before opting for discrimination prevention it is necessary to determine the type of discrimination it may be direct, indirect or combination of both. Depending upon the type of discrimination present in the model following methods can be applied those are:

- A) Pre-processing
- B) In-processing
- C) Post-processing

On the basis of working principle these methods are categorized. Here we will quote the basics of each of the above method.

### A) Pre-processing :

As the name suggests the preprocessing algorithms process the data before actual knowledge extraction i.e. before applying the mining algorithms or the rules to the original dataset. In this method the original dataset is modified in order to achieve the bias free mining rules and finally fair mining model. The principle of preprocessing is such that "The original data is transformed in such way that no biased rule can be inferred from the altered data". The main advantage of preprocessing is that we can apply any mining algorithm to the transformed dataset hence preprocessing technique compatible with any standard data mining algorithms.

The working of preprocessing takes place as follow: The original dataset is grouped into four categories namely DP, DN, PP, PN where first letter of each class represents D stands for Deprived whereas P stands for Privileged and the second letter depicts P for Positive and N for Negative. The next step is to apply the ranker function it sorts the data in ascending order of positive labels later the sample size is changed to make the dataset bias free[11]. Sara Hajian and Josep Domingo-Ferrer proposed an another approach to eliminate the direct and indirect discrimination from the original dataset. They employed 'elift' as the measure for the prevention of discrimination in intrusion detection system [12]. The method of preprocessing for discrimination is applicable in the area where mining is to be performed by third party and data is allowed to publish for the use of public.

Applications: Crime and intrusion detection.

### B) In-processing:

The data present in the original dataset affects the rules and ultimately the mining model. The in-processing methods of discrimination prevention opt to transform the actual algorithms rather than altering the dataset and sacrificing the quality information. The transformation in algorithms eliminates the flaws to generate the fair and bias free mining model. The major distinguishing factor of pre-processing and in-processing is that pre-processing allow any standard mining algorithm to be applied to the dataset and in-processing method requires specially designed algorithms.

The approach proposed by F. Kamiran, T. Calders and M. Pechenizkiy based on decision tree. This approach is further classified into two for the construction of decision aware tree, Dependency Aware Tree Construction and Leaf Relabeling that we discussed in review section of this paper.

The first technique concentrates on splitting criterion for tree construction. Firstly it measures the information gain with respect to class (IGC) and with respect to sensitive attributes (IGS). On the basis of results of basic mathematical operations splitting criteria is decided those are:

(I) IGC - IGS (II) IGC / IGS (III) IGC + IGS. Out of two second method relabels the tree leaves with processing the discrimination aware tree [4].

Applications: Loan approval/denial, Personnel selection, Car Insurance.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

## C) Post-processing :

In previous methods we were either interested in transformation of original dataset or alteration in mining algorithms. These methods are vulnerable to preservation of data i.e. it may result into the loss of quality information if we opt for pre-processing. Sanitizing is not the proper solution to the problem also it is not easy to transform the algorithms in in-processing. Sara Hajian, Anna Monreale, Dino Pedreschi, Josep Domingo-Ferrer proposed post-processing method that modified the final mining model by using protective k-anonymous pattern sanitization to eliminate the discrimination from the data mining model also they derive some frequent classification rules.

The post-processing method is intended to modify the complete mining model instead of data sanitization and altering the algorithms. The disadvantage of post-processing is that it do not allow to publish the data and it is mandatory that mining should be performed by data holder only. T. Calders and S. Verwer proposed Naïve Bayes classifier to remove discrimination from mining model [5].

Applications: Credit assessment, Job recruitment, Insurance computation.

## V. TAXONOMY OF DISCRIMINATION PREVENTION

S. Hajian and J. Domingo-Ferrer described the taxonomy in two dimensions one is the discrimination detection i.e. determining the type of discrimination and the second dimension is to apply the corresponding method to eliminate the discrimination. Following figure depicts the taxonomy of discrimination prevention and its components. Along with applying the discrimination prevention algorithm the discrimination discovery is also important factor and hence there is separate dimension for the discovery of type of discrimination present in the dataset.

The first dimension discovers the discrimination present in the dataset whether it is direct, indirect or the combinations of both, there are attributes on the basis of which we decide the types of the discrimination [7]. The second dimension depicts the method to be employed for the discrimination prevention it may be pre-processing, in-processing or post-processing depending upon the type of discrimination.

The second main aspect is the discrimination prevention figure 2 depicts the necessary steps for generating the bias free dataset or discrimination free data. Figure shows the working model of discrimination prevention model. The process for discrimination prevention starts with the user agent who requests for the data from the original database. As we know that the dataset is biased towards or against certain group and hence this needs to be eradicated from the database in order to fulfil the request of the user fairly.

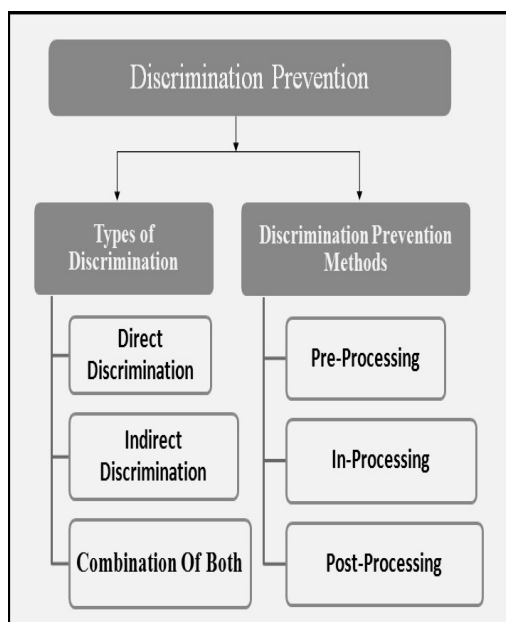


Fig.1 Taxonomy of Discrimination Prevention

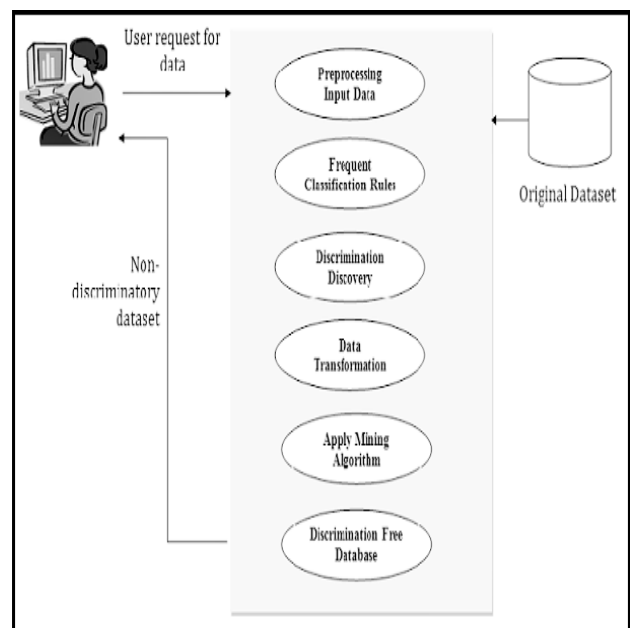


Fig.2 Discrimination Prevention System



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

On the request of the user the detection and prevention system fetches the data from original dataset. Before starting the detection and prevention the data needs to be pre-processed so as to feed as input to the further steps. The next step is that Frequent Classification Rules are applied on the pre-processed data for the purpose of classification among classes and attributes of the classes. The next step we can say is the actual step of working model i.e. Discrimination Discovery here the discrimination present in the hidden dataset is determined for the purpose of applying discrimination prevention technique, it is necessary to detect the type of discrimination present in the dataset. There are several methods of discrimination prevention as we discussed in literature review section of this paper here we are interested in preserving the quality of data along with making it bias free hence here we opted the method of data transformation we did not selected data sanitization or any other method for the same due to the loss of quality information. Once the data is transformed we get the resultant database as bias free, one can go for fair decision making process relying on the dataset we just obtained after data transformation. The next step is to apply any standard or special purpose mining algorithm for extraction of information. The final step is to return the discrimination free dataset to the user for the purpose of automated decision making or for publishing for public usage.

## VI. CONCLUSION AND FUTURE SCOPE

In this paper we surveyed the basic approaches for discrimination detection and prevention in data mining. We precisely discussed that it is evident to eliminate the discrimination present in the automated decision making system. From the survey we can conclude that the pre-processing method is more flexible to apply rather than its other two counterparts the reason behind this is that pre-processing method involves the transformation of biased attributes from the original dataset. While eliminating the discrimination it is important to preserve the quality data and retain the dependency of the original data.

In the paper we noted that decisions are based on historical data or the original dataset. The classification rules are derived from the trained data. In the case that trained data is against certain group then the resulting mining model will be discriminatory it won't make fair decisions.

If the data mining is to be performed by the data holder and the data needs to be published for the public usage then it becomes necessary that data should be discrimination free. Data sanitization i.e. removing discrimination causing attributes from the dataset seems to be simple solution but it results into the loss of quality information also it removes only direct discrimination. It is not guaranteed that discrimination is removed on elimination of sensitive attributes. Some non discriminatory attributes may reveal some information which may result into induction of discrimination in the dataset e.g. Pin code of certain area may reveal the majority of people leaving in the particular area are belongs to certain group and it may result into discrimination.

## REFERENCES

- [1] S. Hajian and J. Domingo-Ferrer, A methodology for direct and indirect discrimination prevention in data mining," Knowledge and Data Engineering, IEEE Transactions on, vol. 25, pp. 1445-1459, July 2013.
- [2] D. Pedreschi, S. Ruggieri, and F. Turini, The discovery of discrimination," in Discrimination and Privacy in the Information Society, pp. 91-108, Springer, 2013.
- [3] F. Kamiran, A. Karim, and X. Zhang, Decision theory for discrimination-aware classification., " in ICDM, pp. 924-929, 2012
- [4] F. Kamiran, T. Calders, and M. Pechenizkiy, Discrimination aware decision tree learning, "in Data Mining (ICDM), 2010 IEEE 10th International Conference on, pp. 869-874, IEEE, 2010.
- [5] T. Calders and S. Verwer, Three naive bayes approaches for discrimination-free classification," Data Mining and Knowledge Discovery, vol. 21, no. 2, pp. 277-292, 2010.
- [6] S. Hajian and J. Domingo-Ferrer, Direct and indirect discrimination prevention methods," in Discrimination and Privacy in the Information Society, pp. 241-254, Springer, 2013.
- [7] Melanie Ann Thomas and Joshila Grace, Prevention of Discrimination in Data Mining," in IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 16, Issue 1, Ver. VII (Feb. 2014), PP 99-101.
- [8] S. Ruggieri, D. Pedreschi, and F. Turini, Dcube: Discrimination discovery in databases," in Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, pp. 1127-1130, ACM, 2010.
- [9] F. Kamiran and T. Calders, Classification with no discrimination by preferential sampling," in Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.
- [10] F. Kamiran and T. Calders, Classifying without discriminating," in Computer, Control and Communication, 2009. IC4 2009. 2nd International Conference on, pp. 1-6, IEEE, 2009.
- [11] Manoj Wakchaure and Shirish Sane, Discrimination Prevention by Different Measures in Direct Rule Protection Algorithm," in International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 9, September 2015.



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. 4, Issue 2, February 2016**

- [12] S. Hajian, J. Domingo-Ferrer, and A. Martnez 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.

## **BIOGRAPHY**

**Pravinkumar S. Chavan, Akshay S. Thakare and Ram P. Tathe** are the UG students of Anuradha Engineering College, Chikhli (MS) India, pursuing their Bachelor of Engineering in computer science and engineering in VIII semester. Their area of interest is Data Mining and its application in daily life.