



*Journal of Advances in  
Science and Technology*

*Vol. III, Issue No. VI,  
August-2012, ISSN 2230-  
9659*

## **A STUDY ON VARIOUS APPROACHES FOR DISCRIMINATION AVOIDANCE IN DATA MINING**

AN  
INTERNATIONALLY  
INDEXED PEER  
REVIEWED &  
REFEREED JOURNAL

# A Study on Various Approaches for Discrimination Avoidance in Data Mining

Dr. Shailendra Singh Sikarwar<sup>1</sup> Mahesh Bansal<sup>2</sup>

<sup>1</sup>Assistant Professor, P. G. V. College, Gwalior

<sup>2</sup>Assistant Professor, P. G. V. College, Gwalior

**Abstract** – Data mining is an important technology for extracting useful patterns from large amount of data. Two major prevalent issues in data mining are privacy violation and discrimination. Discrimination arises when people are given unfair treatment on the basis of their sensitive features like gender, race, religion etc. Types of discrimination are direct and indirect discrimination. Direct discrimination consists of rules based on sensitive attributes like religion, race, community etc. Indirect discrimination occurs when decisions are based on non-sensitive attributes which are closely related to sensitive attributes. Automated data collection and data mining techniques such as classification rule mining are used for making automated decisions by decision support systems. These systems are used for personnel selection, loan granting etc. If the training data sets are biased with respect to the sensitive features, discriminatory decisions may occur. Antidiscrimination techniques including discrimination discovery and prevention have been introduced in data mining. The main purpose of this survey paper is to understand the existing approaches for discrimination prevention.

Automatic data collection has become the most wanted method in the banking sector to make automatic decisions like loan granting/denial. The discriminations in the dataset will lead to take the decisions in the partiality manner. The discrimination can be either direct or indirect discrimination. Direct discrimination occurs when decisions are made based on sensitive attributes. Indirect discrimination occurs when decisions are made based on non-sensitive attributes. To overcome the partiality decisions the proposed system produces the anti-discrimination methodologies. The anti-discrimination methodologies prevent the discriminative decisions in the dataset. The proposed system prevents the discrimination without affecting the data quality.

Data mining is important technology for extracting useful data hidden in large collections of data. Discrimination refers unfair or unequal treatment of people based on membership to a particular category or a minority. Automated data collection and data mining techniques such as classification rule mining have paved way to making automated decisions, like loan granting/denial, insurance premium computation etc. If training data sets are biased regards discriminatory attributes like gender, race etc. discrimination decisions may ensue. For this reason, antidiscrimination techniques including discrimination discovery and prevention have been introduced in data mining. Discrimination can be either direct or indirect.

## INTRODUCTION

Data mining and knowledge discovery in databases are two new research areas that deal with the automatic extraction of useful patterns from large amounts of data. Data mining techniques are used in business and research and are becoming more and more popular with time. There are two issues related to data mining. These issues are privacy violation and potential discrimination.

Discrimination is a very important issue when considering the legal and ethical aspects of data mining. It can be viewed as the act of illegally treating

people on the basis of their belonging to a specific group. People may be discriminated because of their race, ideology, gender, etc, if those attributes are used for making decisions about them like giving them a job, loan, insurance, etc.

Antidiscrimination techniques including discrimination discovery and prevention have been introduced in data mining. Services in the information society allow for automatic and routine collection of large amounts of data. For a given set of information attributes about a customer, an automated system decides whether the customer is to be recommended for a credit or a job selection. Automated such decision systems for

these tasks reduces the workload of the staff of banks and insurance companies, among other organizations. Those data are often used to train association/classification rules in view of making automatic decisions, like loan granting/denial, insurance premium computation, personnel selection, etc. classification rules are actually learned by the system from the training data. If the training data is biased for or against a particular community (e.g., foreigners), the learned model may show a discriminatory behaviour towards that community. The system may interpret that just being foreign is a legitimate reason for loan denial. Find such potential biases and eliminating them from the training data without harming their decision-making utility is therefore important. Data mining must not become a source of discrimination, since automated decision systems learn from data mining models.

Data mining can be both a source of discrimination and a means for discovering discrimination. Types of discrimination are direct and indirect discrimination. Direct discrimination consists of set of rules that are obtained from sensitive discriminatory attributes like gender, religion etc. Indirect discrimination consists of set of rules that are inferred from attributes closely related to the sensitive ones. Beyond discrimination discovery, making knowledge-based decision support systems free from making discriminatory decisions (discrimination prevention) is a more challenging issue. The challenge increases if there is need to prevent not only direct discrimination but also indirect discrimination or both.

There are various approaches available for discrimination prevention in data mining. In order to be able to classify the various approaches, two orthogonal dimensions are used, based on which the existing approaches exist. The first dimension considers whether the approach deals with only direct discrimination, or indirect discrimination, or both at the same time. Based on this dimension, the discrimination prevention approaches are separated into three groups: direct discrimination prevention approaches, indirect discrimination prevention approaches, and direct/indirect discrimination prevention approaches.

Discrimination might be seen as the demonstration of unreasonably treating individuals on the foundation of their fitting in with a particular bunch. Case in point, people may be separated due to their race, belief system, sex, and so forth. In money matters and social sciences, discrimination has been concentrated on for a century. There are numerous choices making assignments which loan themselves to discrimination, e.g. advance allowing and staff choice. In the decades ago, against discrimination laws have been received by numerous fair governments. A few illustrations are the US Equal Pay Act, the UK Sex Discrimination Act, the UK Race Relations Act and the EU Directive 2000/43/ec on Anti-discrimination.

The fundamental commitments of this paper are as takes after: (1) presenting hostile to discrimination in the setting of digital security; (2) proposing another discrimination counteractive action technique dependent upon data conversion that can think about some biased qualities and their fusions; (3) proposing a few measures for assessing the proposed strategy as far as its triumph in discrimination anticipation also its effect on data quality.

Discrimination is termed as the act of unequally treating people on the basis of their belonging to a specific group. For example individuals may be discriminated because of their gender, ethnicity or nationality... etc. Different decision making tasks leads to the discrimination. eg loan granting/denial in the banking application Discrimination is classified into two types. They are direct and indirect discrimination. Direct discrimination occurs when decisions are made based on the sensitive attributes. Indirect discrimination occurs when decisions are made based on the non-sensitive attributes but they are strongly related to the sensitive attributes. To overcome the partiality decisions anti-discrimination method is introduced Anti-discrimination laws have been adopted by many democratic governments. Some examples are The Caste Disabilities Removal Act, Hindu Succession Act 1956, Scheduled Caste and Scheduled Tribe (Prevention of Atrocities) Act. Several data mining techniques have been adapted with the purpose of detecting discriminatory decisions. Anti-discrimination also plays a vital role in the cyber security to detect the intrusion and crime detection.

The main contribution of this paper are as follows (1) Detect the discrimination in the given dataset (2) Prevent the discrimination without affecting the data quality (3) Remove the discrimination by anti-discrimination methodologies and preserve the data quality (4) large amount of data's can be discriminated with the help of anti-discrimination methods Rule protection and rule generalization algorithm are mainly used to generate the discriminate decisions.

## ANTI DISCRIMINATION AND CYBER PROTECTION

In this paper, we use as a running illustration the preparation. It compares to the data gathered by an Internet supplier to discover subscribers potentially going about as gatecrashers. The dataset comprises of nine traits, the last one (Intruder) being the class trait. Every record compares to a subscriber of a telecommunication organization dead set by Subsnun trait. Other than particular traits (Gender, Age, Zip, Race), the dataset likewise incorporates the accompanying traits:

- Downprof: stands for downloading profile and measures the normal amount of data the subscriber downloads month to month. It's

conceivable qualities are High, Typical, Low, Very low.

- P2p: demonstrates if the subscriber makes utilization of peer-to-peer programming, for example emule.
- Portscan: demonstrates if the subscriber makes utilization of a port scanning utility, for example Nmap.

Hostile to discrimination strategies ought to be utilized within the above illustration. In the event that the preparation data are predispositioned towards a certain assembly of clients (e.g. junior individuals), the studied model will indicate unfair conduct towards that bunch and most solicitations from junior individuals will be inaccurately characterized as interlopers.

Moreover, against discrimination strategies could additionally be convenient in the setting of data offering between IDS. Expect that different IDS impart their IDS reports (that hold gatecrasher data) to enhance their individual interloper discovery models. After an IDS allotments its report, this report ought to be disinfected to dodge prompting predisposition biased choices in different IDS.

## ACQUIRING DISCRIMINATION

Discrimination disclosure is about discovering discriminatory choices covered up in a dataset of authentic choice records. The fundamental issue in the examination of discrimination, given a dataset of chronicled choice records, is to quantify the level of discrimination endured by a given bunch (e.g. an ethnic aggregation) in a given connection concerning the order choice (e.g. interloper yes or no).

### 1. Essential Definitions :

- A thing is a characteristic in addition to its worth, e.g. {gender=female}.
- Association/classification manage mining endeavors, given a set of transactions (records), to anticipate the event of a thing dependent upon the events of different things in the transaction.
- An item set is a gathering of one or more things, e.g. {gender=male, Zip=54341}.
- A regular order administer is an arrangement run the show with a backing or certainty more amazing than a specified easier bound. Let DB be a database of unique data records and Frs be the database of regular order runs the show.

### 2. Potentially Discriminatory and Non-Discriminatory Classification Rules :

With the supposition that discriminatory things in DB are decided ahead of time (e.g. Race=black, Age = Young), standards succumb to one of the accompanying two classes concerning discriminatory and non-discriminatory things in DB.

The saying "potentially" implies that a PD standard could presumably accelerate discriminatory choices, so a few measures are required to quantify the discrimination potential. Likewise, a PND guideline could expedite discriminatory choices assuming that joined with some foundation learning, e.g. in the event that in the above illustration one realizes that zip 43700 is basically possessed by dark individuals (aberrant discrimination).

### 3. Discrimination Measures :

Pedreschi et al., interpreted the qualitative proclamations in existing laws, regulations and legitimate cases into quantitative formal partners over arrangement runs the show what's more they presented a group of measures of the degree of discrimination of a PD principle.

The thought here is to assess the discrimination of a administer by the addition of certainty because of the vicinity of the discriminatory things (i.e. An) in the preface of the standard. Without a doubt, elif t is characterized as the proportion of the certainty of the two principles: with and without the discriminatory things. If the principle is to be acknowledged discriminatory can be surveyed by thresholding2 elif t as accompanies.

## LITERATURE REVIEW

Despite the wide deployment of information systems based on data mining technology in decision making, discrimination in data mining did not receive much attention until 2008. Thus, beyond discrimination discovery, a more challenging issue is to prevent knowledge-based decision support systems from making discriminatory decisions. Some of these approaches are related to the discovery and measure of discrimination. The other approaches deal with discrimination prevention.

- A. Classification Rules using  $\alpha$  Protection Measure - D. Pedreschi et al. are the first researchers to address the discrimination problem from the point of view of knowledge discovery from databases. This approach belongs to pre-processing method of discrimination prevention. They have shown that discrimination may be hidden in knowledge discovery models extracted from

databases, and classification rule models are considered.  $\alpha$ -protection is introduced as a measure of the discrimination power of a classification rule containing one or more discriminatory items. The idea is to define such a measure as an estimation of the gain in precision of the rule due to the presence of the discriminatory items. The  $\alpha$  parameter is the key for tuning the desired level of protection against discrimination. This approach is based on mining classification rules and reasoning on them on the basis of quantitative measures of discrimination that formalize legal definitions of discrimination. The limitation of this approach is that it needs to manually check the  $\alpha$ -protection levels.

- B. Discriminatory Classification Rules - S. Ruggieri et al. have introduced the problem of discovering contexts of discriminatory decisions against protected-by-law groups, and provided a knowledge discovery process for solving it. This approach is based on pre-processing method of discrimination prevention. This approach is based on coding the involved concepts (potentially discriminated groups, contexts of discrimination, measures of discrimination, background knowledge, direct and indirect discrimination) in a coherent framework based on item-sets, association rules and classification rules. In direct discrimination, the extracted rules can be directly mined in on basis of discriminatory attributes. In indirect discrimination, some background knowledge is needed by the mining process which when combined with the extracted rules may lead to discriminatory decisions. This approach cannot be applied to continuous attributes.
- C. Three Naive Bayes Approach - T. Calders et al. have presented a modified Naïve Bayes classification approach. This approach belongs to post-processing method of discrimination prevention. The classification task is performed by focusing on independent sensitive attributes. This type of behaviour occurs, when the decision process that leads to the labels in the dataset is biased with respect to sensitive attributes. This approach is motivated by many case studies of decision making, where laws deny a decision that is partly based on discrimination.

Three methods based on Bayesian classifier are used for discrimination-aware classification. In the first method, the observed probabilities in a Naive Bayes model are modified in such a way that its predictions become discriminationfree. The second method involved learning two different model. In the third and most involved method a latent variable  $L$  is introduced reflecting the latent “true” class of an object without discrimination. The This approach is not able to work

with numerical attributes e.g.:- income as a sensitive attribute.

- D. Decision Tree Learning - F. Kamiran, et al have presented the construction of a decision tree classifier without discrimination. This approach belongs to in-processing method of discrimination prevention. They have considered discrimination aware classification as a multi-objective optimization problem.

They have constructed the decision trees with nondiscriminatory constraints. This is a different approach for 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. It outperforms the other discrimination aware techniques by giving much lower discrimination scores and maintaining the accuracy high. This approach is suitable for cases wherein training set is discriminatory and test set is non-discriminatory.

- E. Decision Theory approach - K. Asim et al. have developed two flexible and easy solutions for discrimination-aware classification based on an intuitive hypothesis: discriminatory decisions are often made close to the decision boundary because of decision maker's decisions. Decision theoretic concepts of prediction confidence and ensemble disagreement have been used for this purpose. Their first approach is called Reject Option based Classification (ROC). It makes use of the low confidence region of a single or an ensemble of probabilistic classifiers for discrimination reduction. It invokes the reject option and labels instances belonging to deprived and favoured groups in a manner that reduces discrimination. Second approach is called Discrimination-Aware Ensemble (DAE). It makes use of the disagreement region of a classifier ensemble to relabel deprived and favoured group instances for reduced discrimination. This approach gives better control and interpretability of discrimination-aware classification to decision makers.

- F. Discrimination for Crime and Intrusion Detection - S. Hajian et al. have introduced anti-discrimination in the context of cyber security. They have introduced a new discrimination prevention method based on data transformation that can consider several discriminatory attributes and their combinations. This approach concentrates on producing training data which are free or nearly free from discrimination while preserving their usefulness to detect real intrusion or crime. In order to control discrimination in a dataset, the first step



consists of discovering discrimination. If any discrimination is found, the original training dataset is modified until discrimination is brought below a certain discriminatory threshold or is entirely eradicated. They have introduced some measures for evaluating this method in terms of its success in discrimination prevention and its impact on data quality. The drawback of this approach is that it considers only with direct discrimination.

## CONCLUSIONS

This paper presents a survey of various approaches for discrimination prevention in data mining. From the survey, it can be observed that discrimination prevention is indeed a major issue in data mining. From the survey, it can be observed that approaches based on pre-processing methods are flexible to use than the other two methods since, preprocessing involves transforming dataset so as to remove discriminatory biases from it. The approach is more efficient than the other mentioned approaches since it can handle direct as well as indirect discrimination simultaneously along with preserving data quality.

We have inspected how discrimination could effect on digital security requisitions, particularly Idss. Idss use computational discernment innovations, for example data mining. It is evident that the preparation data of these frameworks could be discriminatory, which might cause them to settle on discriminatory choices when anticipating intrusion or, all the more usually, crime. Our commitment focuses on preparing data which are free on the other hand about free from discrimination while protecting their advantage to identify true intrusion or crime. So as to control discrimination in a dataset, a first stage comprises in running across if there exists discrimination. Assuming that any discrimination is discovered, the dataset will be changed until discrimination is carried underneath a certain limit or is truly wiped out. Sometime later, we need to run our system on genuine datasets, enhance our techniques and additionally think about foundation information (backhanded discrimination).

This paper presents a new pre-processing discrimination prevention method. Different transformations are used for the discovery of discrimination. The process measures the discrimination and identifies the categories by decision-making processes. Discrimination-free data models can be produced from the transformed data set without seriously damaging the data quality. More data's can be handled and the system result is trustworthy.

## REFERENCES

- D.Pedreschi, S. Ruggieri and F. Turini, "Discrimination-aware data mining". Proc. of the 14th ACM International Conference on Knowledge Discovery and Data Mining (KDD 2008), pp. 560-568. ACM, 2008.
- D.Pedreschi, S. Ruggieri and F. Turini, "Integrating induction and deduction for finding evidence of discrimination". Proc. of the 12th ACM International Conference on Artificial Intelligence and Law (ICAIL 2009), pp. 157-166. ACM, 2009.
- D.Pedreschi, S. Ruggieri and F. Turini, "Measuring discrimination in socially-sensitive decision records". Proc. of the 9th SIAM Data Mining Conference (SDM 2009), pp. 581-592. SIAM, 2009.
- 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, ACM, 2008.
- F.Kamiran and T. Calders, "Classification with No Discrimination by Preferential Sampling". Proc. of the 19th Machine Learning conference of Belgium and The Netherlands, 2010.
- F.Kamiran and T. Calders, "Classification without discrimination". Proc. of the 2nd IEEE International Conference on Computer, Control and Communication (IC4 2009). IEEE, 2009.
- F.Kamiran, A. Karim, and X. Zhang, "Decision theory for discrimination-aware classification.", in ICDM, pp. 924-929, 2012.
- 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.
- J.Natwichai, M. E. Orlowska and X. Sun, "Hiding sensitive associative classification rule by data reduction". Advanced Data Mining and Applications (ADMA 2007), LNCS 4632, pp: 310-322. 2007.
- Parliament of the United Kingdom, Sex Discrimination Act, 1975.
- R.Agrawal and R. Srikant, "Fast algorithms for mining association rules in large

databases". Proceedings of the 20<sup>th</sup> International Conference on Very Large Data Bases, pp. 487-499. VLDB, 1994.

- S.Hajian, J. Domingo-Ferrer, and A. Martinez-Balleste, "Discrimination prevention in data mining for intrusion and crime detection," in Computational Intelligence in Cyber Security (CICS), 2011 IEEE Symposium on, pp. 47-54, IEEE, 2011.
- S.R. M. Oliveira and O. R. Zaiane. "A unified framework for protecting sensitive association rules in business collaboration". International Journal of Business Intelligence and Data Mining, 1(3):247-287, 2006.
- S.Ruggieri, D. Pedreschi and F. Turini, "Data mining for discrimination discovery". ACM Transactions on Knowledge Discovery from Data, 4(2) Article 9, ACM, 2010.
- S.Ruggieri, D. Pedreschi and F. Turini, "DCUBE: Discrimination Discovery in Databases". Proc. of the ACM International Conference on Management of Data (SIGMOD 2010), pp. 1127-1130. ACM, 2010.
- S.Ruggieri, D. Pedreschi, and F. Turini, "Data mining for discrimination discovery," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 4, no. 2, p. 9, 2010.
- T.Calders and S. Verwer, "Three naive Bayes approaches for discrimination-free classification", Data Mining and Knowledge Discovery, 21(2):277-292. 2010
- 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.
- V.Verykios, A. Elmagarmid, E. Bertino, Y. Saygin and E. Dasseni, "Association rule hiding". IEEE Trans. on Knowledge and Data Engineering, 16(4):434-447, 2004.
- Y. Saygin, V. Verykios and C. Clifton, "Using unknowns to prevent discovery of association rules". ACM SIGMOD Record, 30(4):45-54, 2001.