

# DATA PREPROCESSING ANALYSIS TO PREVENT DIRECT DISCRIMINATION IN DATA MINING

Miss Megha C. Banapure

PG Student, Department of Computer Science and Engineering, CSMSS, Chh. Shahu  
College of Engineering, Aurangabad, Maharashtra, India  
shubhangimandwale@gmail.com

**Abstract-** Data mining is a technology using which we can extract useful information from data. There are two major issues in data mining first is privacy violation and second is discrimination. Discrimination is the unfair treatment with respect to the features that should not be considered while decision making. With respect to human, it is when people are given unfair treatment on the basis of their sensitive features like gender, race, religion etc. Discrimination can be of two types direct discrimination and indirect discrimination. Direct discrimination consists of training rules based on sensitive attributes like religion, race, community etc. Indirect discrimination is a discrimination which occurs when the decisions are taken on non-sensitive attributes but these attributes are closely related to direct discriminatory attributes. Automated decision making systems uses data mining techniques to train the system for decision making. Data form the previous work is used for the rule generation to train the system. At first sight, we can say that automating decisions systems are fair in decision making, but if the training data sets are itself discriminatory then the system cannot be free from discrimination. To remove such discrimination we have discrimination discovery and prevention techniques in data mining. This paper mainly focuses direct discrimination removal from the data.

**Keywords** –Data Mining, Direct Discrimination, Data Pre-processing, Discrimination Prevention, Rule Protection, Rule Generalization.

## I INTRODUCTION

Discrimination is unlawful act in any means with anyone and so it is punishable act. The manual decision making may or may not cause discrimination depending on the decision maker. As human mind is flexible and can think. But once a system learns it always follow that only. System learn from large amount of data using data mining and knowledge discovery by extracting useful patter from large amount of data. Using the discrimination detection and prevention technique we can prevent law violation. Data mining techniques ate very useful for this. There are two major issues in data mining first is privacy violation and and second is discrimination. Discrimination is an important legal and ethical aspects of data mining. The basis of discrimination is unfair

treatment to people on the basis of there their race, religion, gender, etc. For example using there sensitive information to grant them facilities they deserve.

Data mining can help to stop discrimination. There are two types of discrimination one is direct and other is indirect. Direct discrimination caused when the sensitive information is directly and intentionally use for decision making so as to disadvantage the group in unethical way. Automated systems such loan granting / credit approval can be usefully reduce the staff or manual repeated tasks in the offices. The system learns decision making from the training data. If the training data is itself unfair with respect to some sensitive information then the the system will take discriminatory decisions. For example consider if the training data contain sex discrimination i.e. it do not approve application of female candidate then the system may do so and cause discrimination. And so the data mining will become the source of discrimination. To prevent this we need to clean the training data and make it discrimination free so we are using pre-processing approach.

## II BASIC APPROACH

We can transform discriminatory data into discrimination free by three ways i.e. Pre-processing, In-processing and Post-processing. Pre-Processing is an approach in which the data having discrimination is transformed in such a way that unfair decision rule cannot be mined from the transformed data. Doing so we can apply any of the standard data mining algorithms unlike in-processing where we need to change the data mining algorithm so as the resulting model do not contain discriminatory rules and post-processing in which we need to modify the resulting data mining models, instead of cleaning data set or changing algorithms.

## III LITERATURE SURVEY

### A. Discrimination In Data Mining

Discrimination discovery and discrimination prevention techniques are introduced in data mining to prevent discrimination. Using discrimination discovery technique we can discover that if the dataset using which we are training the automated system is discrimination free or not. As we use the classification rules derived from dataset to train the system. If there are any such potentially biases then we need to eliminate

those from the data that too preserving the quality of data i.e. preserving decision making ability.

**B. Discovering Discrimination**

*Potentially Discriminatory and Non-Discriminatory Classification Rules*

Let DIs be the set of predetermined discriminatory items in database. For example  $DIs = \{ Foreign\ worker = Yes, Race = Black, Gender = Female \}$ .

Frequent classification rules follows into one of the following two classes.

A classification rule  $X \rightarrow C$  is potentially discriminatory (PD) when  $X = A, B$  with A belong to non-empty discriminatory item set and B belong to non-discriminatory item set.

For example :

$\{ Foreign\ worker = Yes, City = NYC \rightarrow Hire = No \}$

A classification rule  $X \rightarrow C$  is potentially non-discriminatory (PND) when X is a non-discriminatory itemset.

For example  $\{ Zip = 10451, City = NYC \rightarrow Hire = No \}$  or  $\{ Experience = Low, City = NYC \rightarrow Hire = No \}$

Here “potentially” means that a Potentially Discriminatory rule that probably lead to discriminatory decisions. Therefore these are the rules which mostly lead to discrimination but sometime may not lead. Same for PND these rules are the rules which generally free from discrimination but sometime they may be the cause of discrimination when combined with some related background knowledge.

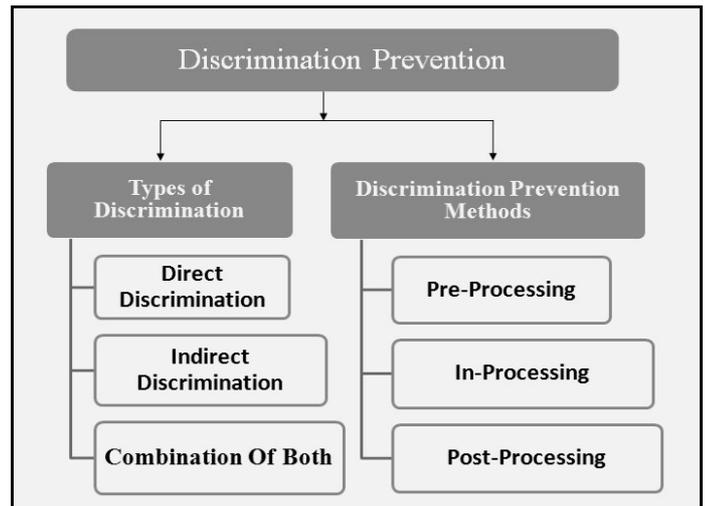
**C. Taxonomy of Discrimination Prevention**

Basically 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.

The second dimension represent 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 thing is the discrimination prevention represent the necessary steps for generating the bias free dataset or discrimination free data.

Figure shows the working model of discrimination prevention. The process for discrimination prevention starts with the user agent who requests for the data from the original database DB. As 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.

On the request of the user the detection and prevention system fetches the data from original dataset. Before starting the detection and prevention in discrimination the data needs to be pre-processed so as to feed as input to the further steps.



**Figure 1: Taxonomy of Discrimination Prevention**

The next step is that Frequent Classification Rules FR 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 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 proposed technique for 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 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.

**D. Direct Rule Generalization**

If each discriminatory rule  $r': A, B \rightarrow C$  in the database of decision rules is an instance of at least one non-redlining (legitimate) PND rule  $r : D, B \rightarrow C$ , the data set would be free of direct discrimination.

In rule generalization, we consider the relation between rules i.e. we compare two rules with each other discriminatory rule is compared with legitimate rule.

If the discriminatory rule is an instance of PND rule then that rule is discrimination free.

Let us take an example of discrimination against foreign workers.

A classification rule

$$\{ \text{Foreign worker} = \text{Yes}, \text{City} = \text{NYC} \rightarrow \text{Hire} = \text{No} \}$$

with high supports discrimination.

However, we can conclude that this rule is an instance of a more general legitimate rule

$$\{ \text{Experience} = \text{Low}, \text{City} = \text{NYC} \rightarrow \text{Hire} = \text{No} \}$$

We can say that, foreign workers are rejected because of their low experience and not because they are foreign workers, and the low experience is very valid reason for rejection

**E. Data mining for discrimination data analysis**

Discrimination discovery from data or information consists in the actual discovery of discriminatory situations and practices hidden in a large amount of decision records. The aim is to unveil contexts of possible discrimination on the basis of legally-grounded measures of the degree of discrimination suffered by protected-by-law groups in such contexts. The legal principle of under-representation has inspired existing approaches for discrimination discovery based on pattern mining in the discrimination.

Data mining propose preprocessing methods which overcome the above limitations and introduces new data transformation methods i.e., rule protection and rule generalization (RG) are based on measures for both direct and indirect discrimination and can deal with several discriminatory items.

**F. PD and PND Classification Rule:**

Assume DC be the arrangement of predetermined discriminatory items in the DB. FP is the frequent classification rule which fall into the accompanying classes:

1. Classification rule  $Y \rightarrow C$  is PD when  $Y = A, B$  with A is the subset of DC a nonempty discriminatory item

The proposed approach for prevention of direct and indirect discrimination is split up into two phases:

set and B is the nondiscriminatory item set.

2. Classification rule  $Y \rightarrow C$  is PND where  $Y = D, B$  is a nondiscriminatory item set.

The expression "potentially" implies a PD manages presumably prompt discriminatory decision. Subsequently, for evaluating the direct discrimination potential a few measures are required.

**G. Pre-processing Approach:**

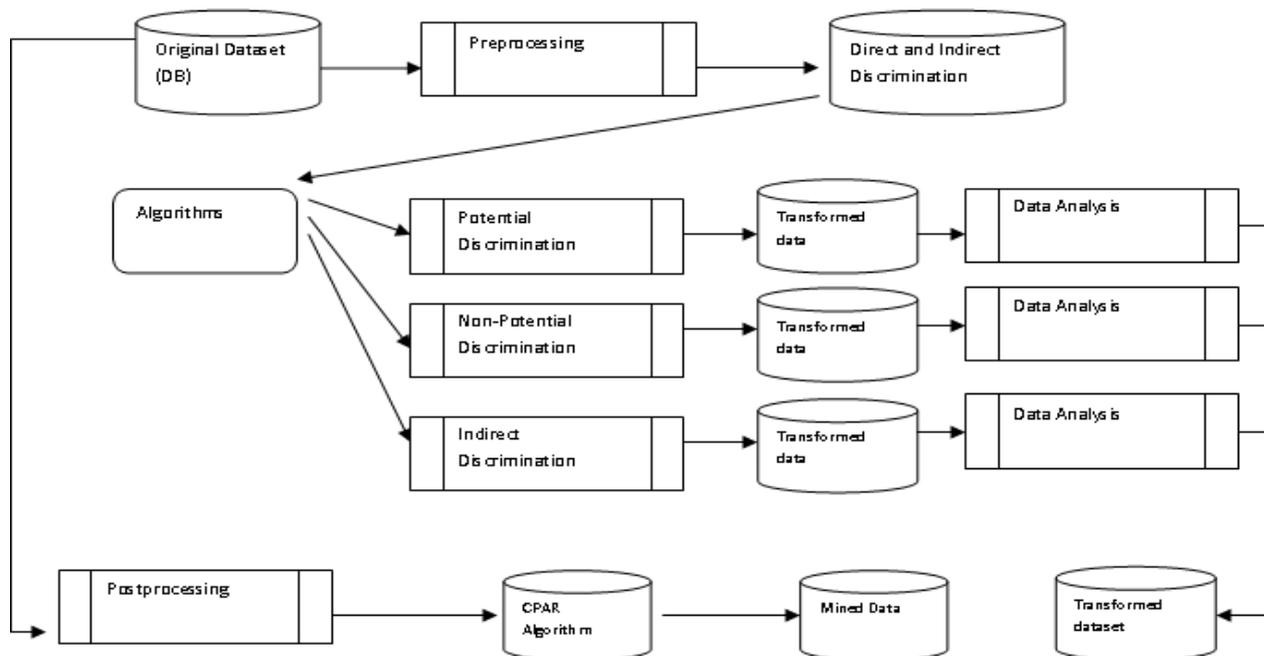
The method of pre-processing approach for prevention of direct and indirect discrimination is split up into two phases:

- Measurement of discrimination:

Direct and indirect discrimination detection contains obtaining the alpha discriminatory rules and redlining rules. At first, PD and PND tenets are created on the premise of unfair things in the database DB and FP the regular classification principle. After that by utilizing the direct discriminatory measures and the oppressive edge the direct separation is measured by getting the alpha discriminatory rules with the PD rules. After that, same as the direct discrimination, indirect discrimination is measured by getting the redlining guidelines with the PND rules joining with the background knowledge, utilizing an indirect discrimination measures and the discrimination threshold.

- Transformation of discrimination:

Changing the original database DB in such path that to remove direct or indirect discriminatory, together with slightest effect on the data and on legitimate decision rules, so that no unjustifiable decision rules can be mined from the exchange database.



**Figure 1 : Proposed approach for prevention of direct and indirect discrimination**

- **Measurement of discrimination**

Direct and indirect discrimination detection contains obtaining the alpha discriminatory rules and redlining rules. Initially, PD and PND rules are generated on the basis of discriminatory items in the database DB and FP the frequent classification rule. After that by using the direct discriminatory measures and the discriminatory threshold the direct discrimination is measured by obtaining the alpha discriminatory rules with the PD rules. After that, same as the direct discrimination, indirect discrimination is measured by obtaining the redlining rules with the PND rules combining with the background knowledge, using an indirect discrimination measures and the discrimination threshold.

- **Transformation of discrimination**

Transforming the original database DB in such way that to eliminate direct or indirect discriminatory, together with least impact on the data and on rightful decision rules, so that no inequitable decision rule can be excavate from the transaction database . We have two methods for direct rule protection given by In this section we discussed about our proposed approached including data transformation methods which used for prevent of direct or indirect discrimination. The proposed system clean the training data sets and the discriminatory rules are converted to legitimate classification rules that too preserving the data quality. To know the effect of the discrimination prevention, we need the utility measures to compare the results before and after.

The previous work done is on direct and indirect discrimination prevention. We are here focusing on only direct discriminatory biases to remove from data in more efficient way .We are going to combine the two direct rule protection methods specified in .

Dataset we are using is Adult dataset available on UCI Repository. The dataset need to be clean to use it so we cleaned it.

*A. Project objective*

- To Discover Discrimination in data
- Transform data in the proper way to remove all direct discriminatory biases
- To Produce direct discrimination free dataset
- To Maintain the data quality
- To Evaluate the measures

*B. Project Outcomes*

- System will be able to make automated decision direct discrimination free.
- System can preserve the privacy of the sensitive data.
- System can take the discrimination free decisions based on legitimate rules.

- System can compare the discrimination prevented and the data quality after direct discrimination removal.

*C. Software Architecture*

1) *Presentation Layer*: Presentation (Desktop) Layer contains the User Interface and is the topmost layer of the application. This layer coordinates between user and system and interface for information exchange. Generally this layer contains forms.

2) *Data Layer*: It is the middle-layer and it has the business service methods, data access logic, data model and SQL scripts.

3) *Data Model*: The data model contains properties that reflect the data; this is often called entity/business objects.

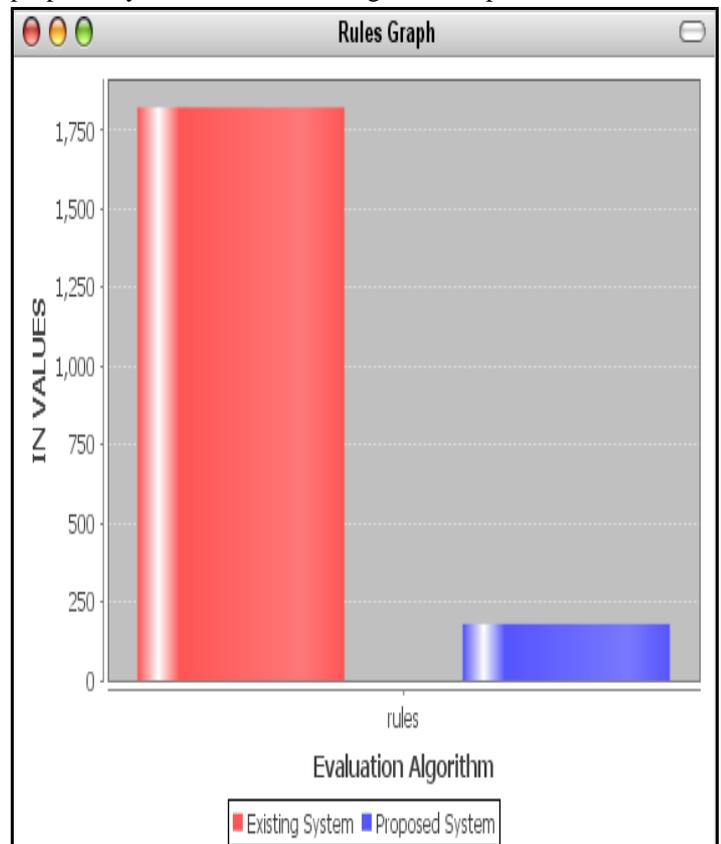
4) *SQL Scripts*: SQL Scripts contains SQL statements that carry the database logic.

5) *Data access*: Data access typically contains the logic for accessing the database data and is generally implemented in a separate class library project. In the demo project Business Service and Data Access are implemented in the same class library.

**IV RESULT & PERFORMANCE ANALYSIS**

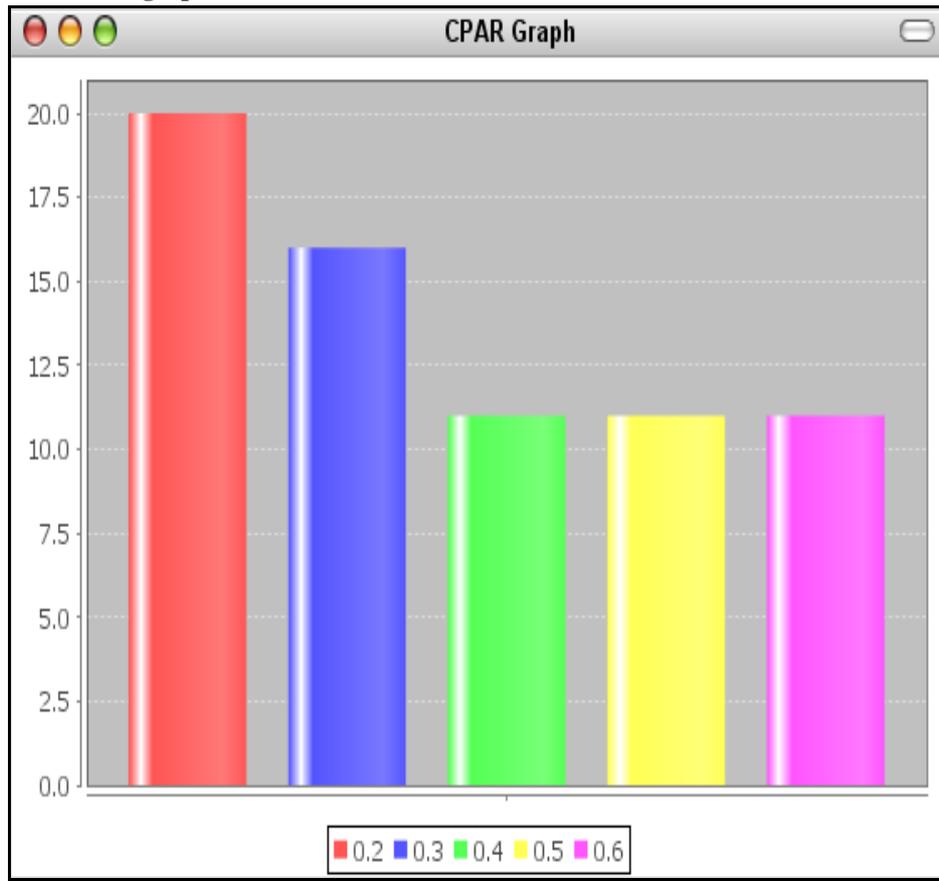
- Comparison of existing system and proposed system (Rule Generation)

Figure represent the comparison of existing system and proposed system in terms of rule generation process



**Figure 2: Rules Graph**

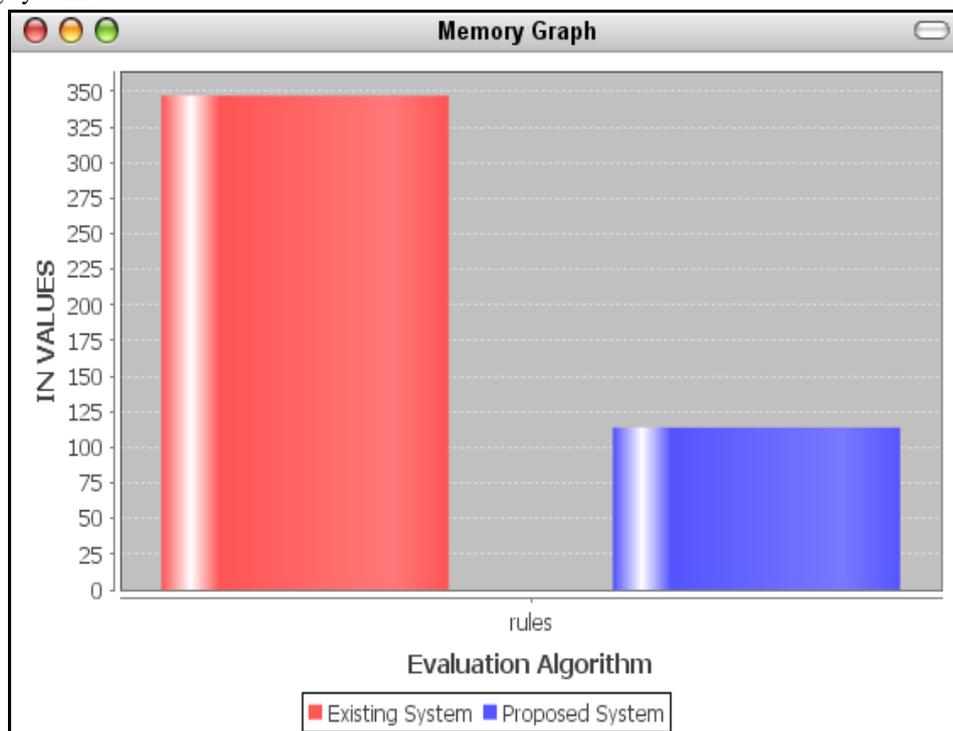
• **CPAR Algorithm rules graph**



**Figure 3: CPAR Rules Graph**

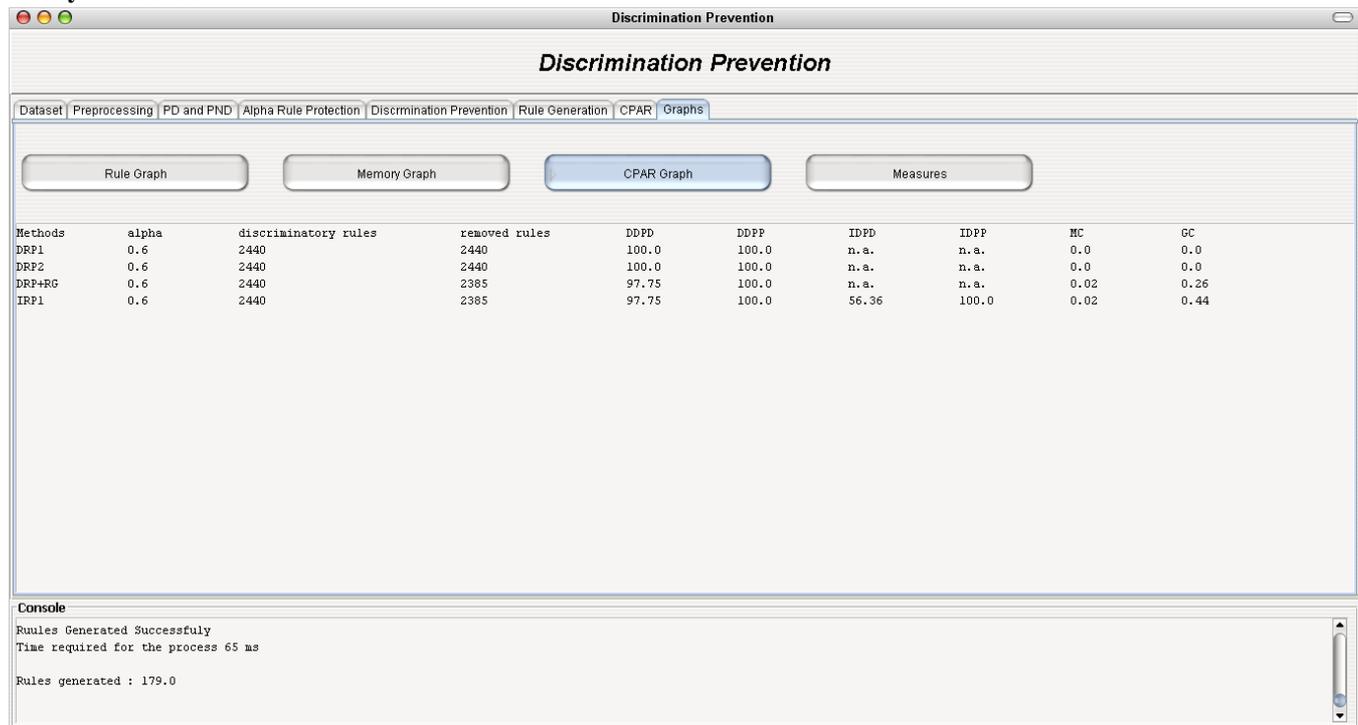
• **Comparison existing system and proposed system (memory Utilization)**

Represent the comparison of existing system and proposed system in terms of memory utilization. Proposed system utilizes less memory than existing system.



**Figure 4: Memory Comparison**

**Utility Measures:**



**Figure 5: Utility Measures**

**V CONCLUSION**

Discrimination prevention is a major issue in data mining. It is studied that approaches based on pre-processing methods are flexible to use than the other two methods in pre-processing and post processing since; preprocessing involves transforming dataset so as to remove discriminatory biases from it. The study of direct discrimination removal include the discrimination detection, rule generalization and discrimination prevention. Studying this we are proposing a technique which will prevent the discrimination more precisely as we are going to combine both method of direct rule protection and remove the discrimination.

**REFERENCES**

[1] Hajian, Sara, and Josep Domingo-Ferrer. "A methodology for direct and indirect discrimination prevention in data mining." *Knowledge and Data Engineering, IEEE Transactions on* 25.7 (2013): 1445-1459.

[2] Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." *Proc. 20th int. conf. very large data bases, VLDB. Vol. 1215.* 1994.

[3] Calders, Toon, and Sicco Verwer. "Three naive Bayes approaches for discrimination-free classification." *Data Mining and Knowledge Discovery* 21.2 (2010): 277-292.

[4] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[5] Hajian, Sara, Josep Domingo-Ferrer, and Antoni Martinez-Balleste. "Discrimination prevention in data mining for intrusion and crime detection." *Computational Intelligence in Cyber Security (CICS), 2011 IEEE Symposium on.* IEEE, 2011.

[6] Pedreshi, Dino, Salvatore Ruggieri, and Franco Turini. "Discrimination-aware data mining." *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2008.

[7] Ruggieri, Salvatore, Dino Pedreschi, and Franco Turini. "Data mining for discrimination discovery." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 4.2 (2010): 9.

[8] Hajian, Sara, and Josep Domingo-Ferrer. "Direct and indirect discrimination prevention methods." *Discrimination and Privacy in the Information Society.* Springer Berlin Heidelberg, 2013. 241-254