Received: February 16, 2018

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An Efficient Discrimination Prevention and Rule Protection Algorithms Avoid Direct and Indirect Data Discrimination in Web Mining

Mylam Chinnappan Babu[1]* **Sankaralingam Pushpa[1]**

[1]*Department of Computer Science and Engineering, St. Peter's Institute of Higher Education and Research, Chennai, Tamilnadu, India*
*Corresponding author's Email: mcbabu.phd@gmail.com

Abstract: Discrimination is one of the most important challenging tasks in web mining due to its many legal and ethical features in social media and enterprise based industries. There are an enormous amount of anti-discrimination measures available to prevent discrimination such as using some features like race, religion, gender, nationality, disability, marital status, and age besides situations like employment and training, access to public services, credit, insurance, etc. Practically, those systems are not possible to use in industries due to large datasets. Indirect discrimination contains a set of rules or techniques which are not explicitly specifying discriminatory features, deliberately or accidentally and could create unfair decisions. Existing systems have low classification accuracy and data loss with high discrimination data detection time. To overcome these limitations, an Efficient Discrimination Prevention and Rule Protection (EDPRP) approach has been proposed for removing the discrimination and protects the rule without damaging the data quality. The proposed system designing pre-processing discrimination prevention approach and specify the different features and represent to deal with direct or indirect discrimination. EDPRP is capable of preventing Indirect and direct discrimination, and it allows automatic and routine collection of large amounts of data from the public. In EDPRP, the discrimination prevention model is based on partial data sets as part of the automated decision making. Based on Experimental evaluations, proposed method improves 8% (percentages) of Support and 8 ms (milliseconds) of Execution Time compared than existing methods.

Keywords: Discrimination, Anti-discrimination, Rule protection, Rule generation, Support, Confidence.

1. Introduction

In sociology, discrimination is the prejudicial treatment of an individual based on their candidature in a particular group or category. It involves rejecting one group of member opportunities which are present in some other groups. There are enormous amounts of anti-discrimination available to prevent discrimination by using some features like race, religion, gender, nationality, disability, marital status, and age besides various situations such as employment and training, access to public services, credit, insurance, etc. Here, discrimination can happen either directly or indirectly. Direct discrimination contains a set of rules or procedures that explicitly mention minority or disadvantaged groups based on sensitive discriminatory attributes. Indirect discrimination contains a set of rules or procedures which are not explicitly mentioning discriminatory attributes that could generate discriminatory decisions.

At first view, automating decisions may provide a sense of fairness. However, their classification rule does not direct themselves based on personal preferences. However, at a closer look realized classification rules are learned by the system from the training data. Current discrimination discovery method considers each rule individually for measuring discrimination without considering other rules or the relation between them. However, the system is not flexible for discrimination discovery, based on the existence or nonexistence of discriminatory data. Here, these systems work individually for indirect discrimination and direct discrimination prevention which can be utilized easily in an industry sector. Existing systems have low classification accuracy with high discrimination data detection time. Technology can be utilized proactively in legislation by participating discrimination discovery work and as well to find some prevention approaches. These data can often be used to train association/classification rules to make automated decisions.

To overcome this limitation, an Efficient Discrimination Prevention and Rule Protection (EDPRP) approach has been proposed to reduce redlining rules. The rules that cause indirect discrimination and be called redlining rules. EDPRP is capable of preventing Indirect and direct discrimination. Even though, it is effective if features have background knowledge (rules), EDPRP allows automatic and routine collection of large amounts of data from the public. In EDPRP, the discrimination prevention model based on partial data sets as part of the automated decision making. It can be utilized with several discriminatory items to detect the discriminations. The research work focuses on discrimination prevention based on preprocessing because the preprocessing approach seems the most flexible. It does not require changing the standard data mining algorithms, unlike in processing approach and it also allows data publishing. It also provides utility measure. Hence, it can be said that the proposed approach to discrimination prevention is broader than the existing work. In this phase, the system first measures discrimination and identify the group of data attributes which are directly or indirectly involved in direct or indirection discrimination process. Hence, it moves for data transformation with proper data classification to remove all those discriminatory biases. The system produces discrimination-free data models to transform data with classification and little discrimination data detection time without reduction in data quality. The rest of paper contribution is followed as:

- To design pre-processing discrimination prevention approach and specify the different features of each approach and represent this approach to deal with direct or indirect discrimination.
- To calculate the discrimination and identify categories and groups of individuals in the decision-making.
- To develop discrimination-free data models that can be produced from the transformed data set without severely damaging data quality
- To improve the discrimination detection accuracy, support, confidence and reduce the execution time compared to existing approaches.

The rest of the paper is organized as Section 2 addresses the reviews the closest work of rule discrimination and protection. Section 3 explains the implementation procedure of proposed methodology and algorithm details. Section 4 evaluates the result of the proposed methodology and discusses their performance details. Section 5 concludes the overall work with the scope future work.

2. Related Work

In [1] designed exploratory discriminate aware data mining (DADM) and the relative merits of constraint-oriented from the conceptual viewpoint. The discrimination-aware tool support in the exploratory eDADM and constraint oriented cDADM treatments led to significantly higher proportions of correct decisions. However, it sometimes takes wrong decisions and low accuracy. In [2] discussed the role of income inequalities and backed by a theoretical model of indirect price discrimination. But, indirect price discrimination is not only probable, but it is not familiar on auto markets. In [3] described a classification method which was called as Discrimination-Aware Association Rule classifier (DAAR). It integrated a discrimination-aware measurement and an association rule mining algorithm. A prediction rule utilizing sensitive aspects may accomplish high accuracy, but it is not acceptable as it is discriminating, which is both immoral and against the law. In [4] introduced the first generalization-based approach which offered privacy preservation and discrimination prevention simultaneously. The authors defined the problem, give an optimal algorithm to tackle it and evaluate the algorithm regarding both general and specific data analysis metrics. However, it fails to maintain privacy preservation and discrimination prevention simultaneously. In [5] designed to protect the privacy of the users from a web search engine that tries to protect them. The system provided a distorted user profile to the internet search engine because some of each user's queries are submitted by his/her friends in the social network. However, it fails to provide end-to-end protection.

In [6] studied that how to generate decision supports models automatically and exhibit discriminatory behavior on behalf of particular groups based on gender or ethnicity. Surprisingly, such behaviour may even be observed when sensitive information removed or suppressed and neutral arguments guide the whole procedure. However, it is not simple to compute and assess such probabilities for indirect discrimination in practical cases. In [7] designed genetic algorithm for automatically generating classification rules from the history of the dataset. In the system, the fitness function is calculated in many ways. The name itself, implies new rules can be generated from history and create better matches for classification rules through several steps. But, it fails to maintain the generating new rules and fitness function. In [8] designed antidiscrimination laws for discrimination prevention. Discrimination may be direct or indirect. Direct discrimination found when decisions made according to sensitive attributes. Indirect discrimination is observed when decisions were set according to non-sensitive attributes which were strongly related to biased sensitive ones. However, it includes unwanted dependencies among sensitive and non-sensitive attributes. In [9] worked on discrimination prevention concept in banking where user's loan application will approve or denied based on discrimination. It focused preventing classification rules which used to find direct and indirect discriminatory attributes with the help of apriori algorithm concept. But, the consequential method does not acquire suitable decision rules in processing approach. In [10] focused on how to clean training data sets and outsourced datasets, such that no discrimination should occur. Since, the discrimination laws (rules) or procedures are not clearly declaring discriminatory features.

In [11] designed an association rules which were protected sensitive items for ensuring safety from privacy threats. The work deals with using privacy preservation technique to improve the discrimination prevention system. However, it fails to maintain discrimination prevention methods for sensitive items. In [12] studied alternative approach for association rules mining to enhance the Apriori algorithm and reduce its time complexity. But, it is not good method for huge database and it permitted only minimum support threshold. In [13] reviewed the existing methods for discrimination prevention and analysis of existing approaches to discrimination prevention. But, it cannot handle indirect discrimination. In [14] designed averaging approach using fuzzy logic for contributing the data in different clusters and introduce the user more relevant threshold automatically. Since, it failed to determine a suitable threshold for averaging algorithm. In [15] designed statistical method to analyze the quality rule of the apriori algorithm in association rule mining for splitting the interesting rules within massive association rules. FP-Growth permitted frequent itemset detection without candidate itemset generation algorithm. Eclat was an effective algorithm for mining all the frequent itemsets. However, it fails to maintain quality of rules and frequent itemsets.

In [16] developed a method to measure frequency rates of cyber grooming, profiled characteristics of cyber grooming perpetrators, and examine direct and indirect associations between cyberbullying victimization. But, it also considered risk factor and effects of abuse. In [17] implemented the system that provides discrimination prevention as well as privacy preservation with classification and clustering of the data. However, the discrimination elimination in privacy preserving data mining is not estimated and does not concentrate the quantity of information loss. In [18] reviewed the latest existing on behalf of antidiscrimination techniques and also focuses on discrimination discovery and prevention in data mining. But, it does not a run approach on real datasets and does not consider background knowledge (indirect discrimination). In [19] designed the multilevel privacy preserved anti-discrimination method for free data transmission and deals with the correlation of discrimination prevention. Since, it is inadequate to prevent aspect disclosure and cannot concentrate the information loss. In [20] implemented multi-objective optimization (EMO) algorithm to the tradeoff between sensitive hiding rules and disclosing non-sensitive ones during hiding process for finding a suitable subset of transactions. However, it is hard to hide all sensitive and non-sensitive rules without any side effects. In [22] designed Boolean association rules based on the width preference-traversing manner and it depends on support and certainty, which is provided through the width preference-traversing manner. However, it does not improve the framework of the association rules algorithm depends on the support and confidence. In [23] developed multicriteria decision-making approach depends on ELECTRE methodology. It selected the most utilized association rules produced utilizing apriori. But, it fails to maintain association rules with apriori.

3. System methodology

This system expresses the implementation details of proposed system methodology in detail along with algorithm explanation. The system designs pre-processing Effective Discrimination Prevention and Rule Protection approach to

Received: February 16, 2018

specifying the different features of each approach and represents this approach to deal with direct or indirect discrimination. EDPRP is classified in following modules like user, administrator, manager, rule generation, rule protection, direct discrimination prevention, indirect discrimination, Effective Discrimination Prevention and Rule Protection Algorithm. The systematic workflow of the proposed system is explained in figure (refer with: Fig. 1) in details.

3.1 Rule protection component

The data transformation is carried out to provide direct and indirect rule protection. Classification rules do not provide themselves with individual preferences. However, the classification rules are essentially trained by the system from the training data. The training data are essentially biased for or against a particular community (e.g., foreigners); may produce discriminatory rules.

3.2 Rule generalization

The data transformation is carried out to perform on direct rule generalization and indirect rule generalization. In the rule generalization, it considers the relation between rules instead of discrimination evaluation. The general rule filtering minimum-experienced candidates is a legitimate one because experience can be considered a genuine/legitimate requirement for some jobs.

3.3 Direct discrimination prevention

Direct discrimination is found when decisions are made based on sensitive attributes. It contains a set of rules or procedures which explicitly mention minority or disadvantaged groups based on a sensitive discriminatory attribute associated with group membership. To avoid the issues of the fact, decision rules would be free from direct discrimination. The apply rule protection and rule generalization in this context.

3.4 Indirect discrimination

Indirect discrimination occurs once decisions are made based on non-sensitive attributes which are strongly correlated with biased sensitive ones. It consists of rules or procedures that, while not explicitly mentioning discriminatory attributes, could generate discriminatory decisions. To prevent indirect discrimination it is based on the fact that the data set of decision rules would be free of indirect discrimination if it does not contain redlining rules.

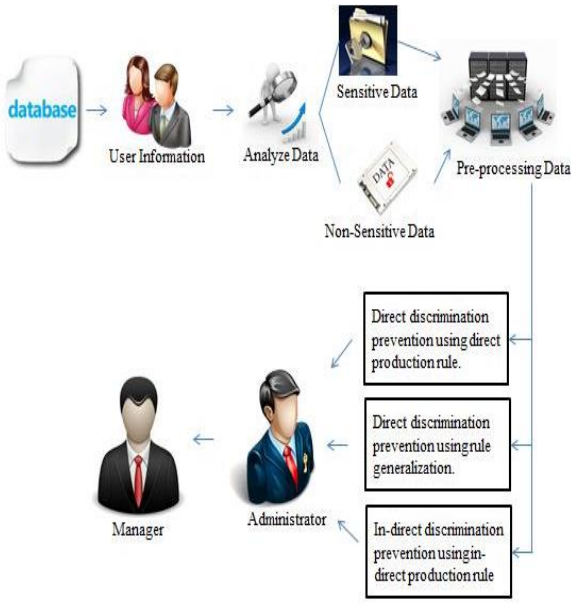

Figure. 1 Workflow of proposed system

Redlining rules designate biased rules that are indirectly assumed from non-discriminatory items owing to their correlation with discriminatory ones. It collects this information; flexible data contribution with minimum information loss is applied for redlining the rules which are converted to non-redlining rules.

3.5 Effective discrimination prevention and rule protection algorithm

The proposed Effective Discrimination Prevention and Rule Protection (EDPRP) Algorithm is to measure direct as well as indirect discrimination detection and prevention without compromising on data quality. The proposed work in this section technique prevents indirect discrimination and is based on the fact that the dataset design set of decision rules to free the indirect discrimination from data if it does not have redlining rules. To achieve good data transformation scalability, this approach is applied in redlining rules which are converted to non-redlining rules. This system resolves the information loss problem which happened during data transformation at the time of direct and indirect discrimination. The technique conducts the process called as a discrimination measurement to discover the direct and indirect discrimination detection. Next, it applies data transformation to remove direct and indirect discriminatory biases with minimum impact on the data with legitimate decision rules. It can be used to provide direct data rule and indirect data rule privacy at the same time during data transformation from one point to another point. The technique also

 DOI: 10.22266/ijies2018.0831.21

takes cares of the relation between item and dataset. This system maintains high-level classification accuracy with minimal discrimination detection time and low computational cost. The algorithm is divided into following steps: Pre-processing: the source data to reduce discriminatory biases. In-processing: Alter the data mining algorithm to avoid the unfair decision rules from resulting models. Post-processing: The method modifies the resulting data mining models.

A data set is a collection of data objects (records) and their attributes. Let DB be the original data set. An item is an attribute along with its value, e.g., nationality = Indian. An item set: i.e., X is a collection of one or more items, e.g., Experienced Worker = Yes; City = Sydney. This system attribute wise spilt the data. Hence, it evaluates the ratio of each attribute REA.

This design maps the attribute in DB to make a relationship with item X. Efficient Discrimination Prevention and Rule Protection (EDPRP) detects the discriminated attribute (DA) and measure the support and confidence factor of each attribute. The pseudo code of proposed EDPRP algorithm given below:

Input: Load the input Dataset DB (German Credit card data sets)
Output: Visualize the Tabular Result TR with quality of Information (support, confidence, execution time and information loss)

Process:
Load the DB;
Identify the Attribute of Item X;
Item→X is the ideal classifier for item
DB= $\{(item_1, x\ (item_1)), \ldots .(item_n, x(item_n))\} \subseteq$ item x X is a set of examples.
Spilt the REA
Splitting based on attribute with domain $\{a_1, \ldots .a_n\}$
REA= $\{item \in X: item/Attribute=a_1\} U \ldots . U\{ item \in X: item/Attribute=a_k\}$
If $\forall$ $\{(item, X(item))\} \in DB:X(item)=X$ Then
Return (item)
End if
If Attribute= $\emptyset$ then
Return (item)
End if
Apply EDPRP;
Evaluate the ratio of AW;
AW=$argmax_{A \in Attribute}$ (AttributeGain(DB,A));
Apply the tree based classification on AW;
For each $\alpha \in$ AW do
DB_a= $\{(item, X\ (item) \in DB: item \mid Attribute=\alpha)\}$
If $D_a=\emptyset$ then
Item=most common class (DB, AW)
End if
Measure the Support and confidence of AW;
Support = Frequency of occurrences that contain both items and attribute
Confidence = Measures how often items in Attribute appear in occurrences that contain items
Generate the rule for DB;
Maps the relationship of AW in DB;
Identify the fraction records of AW in DB
If DA is detected then
Apply K_NN classifier;
Apply Attribute_Selection_Method (DB, Attribute List)
Split attributes value (AW, Attribute List)
If splitting attribute is discrimination value
AW satisfies the outcome value
Prevent the Discrimination of AW
Calculate Information loss values
Else
AW discrimination is not prevented
Else
Tabular Result TR with quality of Information;
Evaluate the performance;
End

4. Result and discussion

4.1 Implementation setup

The implementation is deployed in Intel Dual Core Processor with 1GB RAM running with a windows7 ultimate laptop. The developed is done in JAVA with NetBeans 8.0.2, Apache Tomcat 8.0.15, MYSQL 5.5 databases and Weka library. The proposed method evaluated with German Credit card datasets.

4.2 Performance matrix

The mathematical expression of the proposed algorithm is elaborated to evaluate the system performance on respective parameters. The following parameters are used to find the accuracy and improve the quality of data.

4.3 Support

Support represents the frequency of the rule within transactions. A high value means that the rule involves a great part of a database. The proposed methodology is defined mathematical model for Support in Eq. (1). Support is computed as:

$$S = \frac{\sigma(X \curlyvee Y)}{\#ofTransaction} \quad (1)$$

Where σ frequency of occurrence of an item is set X and Y, and # of the transaction is the total transaction of items set.

4.4 Confidence

The confidence expressed as the percentage of transactions containing A which also contain B. It is an estimation of conditioned probability. The proposed method is described mathematical model for Confidence in Eq. (2). Confidence is estimated as:

$$C = \frac{\sigma(X \cup Y)}{\sigma(X)} \tag{2}$$

Where σ frequency of occurrence of the item is set X and Y.

4.5 Information loss

The Information Loss Ratio is defined as the trade-off between information gain (IG) and privacy loss (PL). The proposed method is defined mathematical model for Information Loss in Eq. (3). Information loss is calculated as:

$$Information\ Loss = \frac{Information\ Gain(IG)}{Privacy\ loss\ (PL)} \tag{3}$$

$$IG = E(T[c]) - \sum_c \frac{|T[c]|}{|T[v]|} E(T[c]) \tag{4}$$

T[v] is the T set of records Generalized Value of v, E (T[c]) is less entropy and T[c] is the T is set of records Child Value of v. The Information gain is described mathematical model in Eq. (4).

$$PL = avg\{A(QID_j) - A_s(QID_j)\} \tag{5}$$

Where $A(QID_j)$ and $A_s(QID_j)$ describe the anonymity of QID_j before and after the specialization. The privacy loss is defined mathematical model in Eq. (5).

4.6 German credit card datasets

The dataset comprises of 1,000 records and 20 attributes (without class attribute) of financial balance holders. It is an excellent real-life data set, including both numerical and categorical attributes. It has frequently been utilized as a part of the antidiscrimination. The class attribute in the German Credit dataset values representing the good or bad classification of the financial balance holders. For our examinations with this data set, we set DBs = {Foreign worker = Yes, Personal Status =Female and not Single, Age = Old}; (cut-off for Age = Old: 50 years old). Direct discrimination is removed without reducing the data quality. Table (refer with: Table 1) shows estimation outcomes of the different techniques and also the comparison between them.

Table 1. Represents the support, confidence for German credit card data sets

Learning Algorithm	Support	Confidence	Execution Time
Apriori	0.2	1	115
FP-GROWTH	0.42	0.43	179
Eclat	0.59	0.92	65
EDPRP	0.67	1	57

Table 2. Discrimination prevention detection measures

Methods	Prevention Measures ($PM = \frac{\lvert PR \rvert \cap \lvert PR` \rvert}{\lvert PR \rvert}$)
Direct Discrimination	0.9091
Direct Rule Protection	1.0005
Indirect Rule Protection	2.4350
Rule Generalization	19.528

Table (refer with: Table 1) illustrates Support (S), Confidence (C) and Execution Time (ET) for German credit card dataset. The proposed EDPRP system is computed with following existing methods namely Apriori [21], FP-GROWTH [21] and Eclat (Equivalence CLASS Transformation) [21] methods. The proposed EDPRP is improving the discrimination detection, support, confidence and reduces the execution time. It noticed that EDPRP method has the best score on every particular constraint for respective parameter.

Table (refer with: Table 2) demonstrates the discrimination prevention measures result in detail. The result of direct discrimination, direct rule protection, indirect rule protection and rule generalization details are given in the table. It claims that proposed system performs the best result in direct discrimination, direct rule protection, indirect rule protection and rule generalization in comparison with the existing methods.

Where PR is the database of α-protective rules extracted from the original data set DB and PR' is the database of α- protective rules extracted from the transformed data set DB' [24].

Table (refer with: Table 3) expresses the discrimination prevention measurement with threshold that shows the discrimination prevention ratio. The threshold value is called α. Its represent the accuracy of information & α value is evaluated on 0.1 to 0.10 to evaluate the information losses [24]

Received: February 16, 2018

which details are given in Table (refer with: Table 3).

Where MR is the database of α-discriminatory rules from DB and MR' is the database of α-discriminatory rules extracted from the transformed data set DB'.

Table 3. Discrimination prevention measures

α - Value	Discrimination prevention Measurement $DPM = \frac{\vert MR\vert - \vert MR`\vert}{\vert MR\vert}$
0.1	0.9096
0.2	0.8338
0.3	0.7696
0.4	0.7147
0.5	0.6670
0.6	0.6253
0.7	0.5885
0.8	0.5558
0.9	0.5266
0.10	0.9096

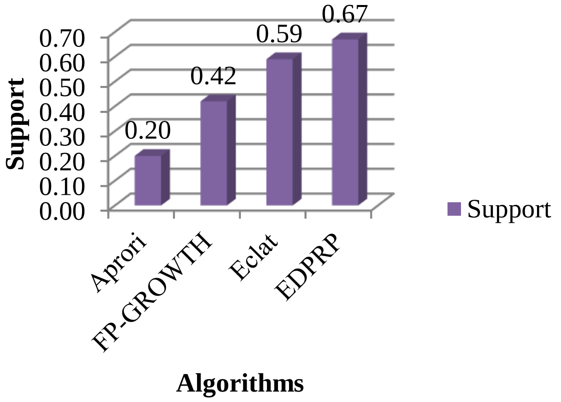

Figure. 2 Support (0-1) for German credit card data sets

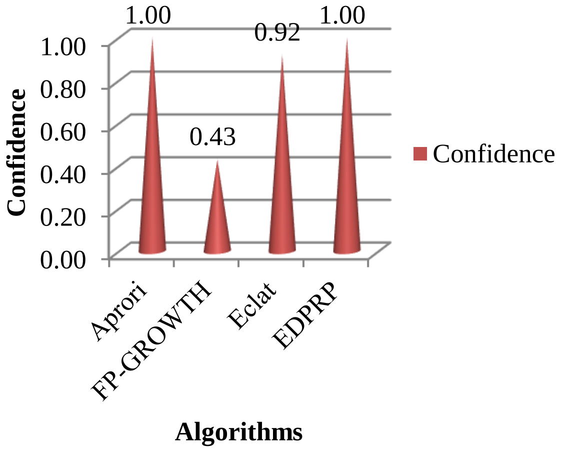

Figure. 3 Confidence (0-1) for German credit card data sets

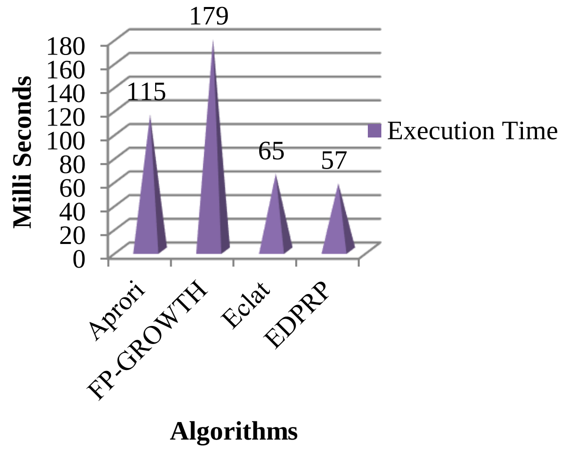

Figure. 4 Execution time (MS) for German credit card data sets

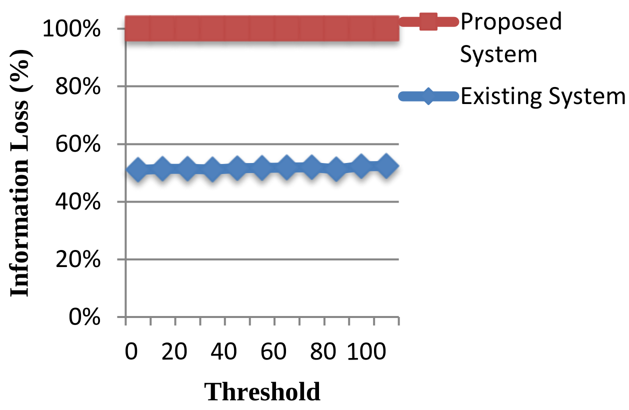

Figure. 5 Information loss ratio of proposed method

Based on Figure (refer with: Figs. 2, 3, 4, and 5), it is observed that the proposed Efficient Discrimination Prevention and Rule Protection (EDPRP) performed better on respective parameters namely support, confidence and execution time. The apriori approach is utilized to splitting the interesting rules within massive association rules. However, it fails to maintain analysis of quality rules. The EDPRP is maintaining quality of rules effectively. The FP-growth algorithm is permitting frequent itemset detection without candidate itemset generation. But, it fails to maintain the candidate itemset generation. The EDPRP is effectively maintains the candidate set generation. Regarding support and execution time, Eclat is the closest method to proposed EDPRP approach. Eclat algorithm predicts frequent item sets in a transaction from large amount of data set. However, it fails maintains frequent item set and take more time to predict frequent item set. Regarding confidence proposed EDPRP outperformed by the Apriori algorithm. However, proposed EDPRP algorithm performs best result for support and execution time.

Figure (refer with: Fig. 5) expresses the information loss ratio with non-discriminated data. The non-discriminated data is predicted by KNN (K-Nearest Neighbour) classifier from dataset to achieve the best accuracy. The proposed EDPRP method improves 8% of support and minimizes 8ms of execution time. Finally, it can be said that the proposed EDPRP is the best approaches for all dataset on respective parameters.

5. Conclusion

The paper proposed and implemented Efficient Discrimination Prevention and Rule Protection (EDPRP) algorithm to measure direct and as well as indirect discrimination detection and prevention without reducing the data quality. This technique conducts the process called as a discrimination measurement to discover the direct and indirect discrimination detection and the second one is data transformation to remove direct and indirect discriminatory biases with minimum impact on the data and legitimate decision rules. This work focuses on discrimination prevention based on pre-processing because the pre-processing approach seems the most flexible. It does not require changing the standard data mining algorithms, unlike the processing approach, and it also allows data publishing. This system maintains high-level classification accuracy with minimal discrimination detection time and low computational cost. EDPRP produced the best result, for support and execution time, and the closest approach was Charm algorithm. Regarding confidence proposed EDPRP is outperformed by the CFP-Growth algorithm. However, CFP-Growth execution time and support performance is very low compared to EDPRP improves 8% support and reduces execution time 8ms.

In future, this work can be extended to work on the distributed dataset in a cloud environment where privacy preserving of sensitive data is a challenging work.

References

[1] B. Berendt and S. Preibusch, "The better decision supports through exploratory Discrimination-aware data mining: foundations and empirical evidence", *Artificial Intelligence and Law*, Vol. 22, No. 2, pp. 175-209, 2014.

[2] V. Radoias, "Direct and indirect price discrimination in the automotive industries of the European Union", *Empirical Economics*, Vol. 50, No. 3, pp. 975-990, 2016.

[3] L. Luo, W. Liu, I. Koprinska, and F. Chen, "Discrimination-Aware Association Rule Mining for Unbiased Data Analytics", In: *Proc. of the International Conference on Big Data Analytics and Knowledge Discovery*, pp. 108-120, 2015.

[4] S. Hajian, J. Domingo-Ferrer, and O. Farràs, "Generalization-based privacy preservation and discrimination prevention in data publishing and mining", *Data Mining and Knowledge Discovery*, Vol. 28, No. 5-6, pp. 1158-1188, 2014.

[5] J. A. Manjón and J. Domingo-Ferrer, "Selected Privacy Research Topics in the ARES Project: An Overview", In *Advanced Research in Data Privacy*, pp. 15-25, 2015.

[6] T. Calders and L. Žliobaitė, "Why unbiased computational processes can lead to discriminative decision procedures", *Discrimination and Privacy in the Information Society*. Springer Berlin Heidelberg, pp. 43-57, 2013.

[7] A. Saheeda, K. A. Shanavas, and G. Sandra, "A Method Based on Genetic Algorithm for Discrimination Discovery", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 4, No. 7, pp.127-132, 2014,

[8] N. T. Qazi and V. M. Deshmukh, "A Methodology for Sensitive Attribute Discrimination Prevention in Data Mining", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 3, No. 5, pp. 3953-3958, 2015,

[9] P. Meshram and S. Bodkhe, "A Result on Privacy Preservation Method by Applying Discrimination Rules in Data Mining", *International Journal of Science, Technology & Management*, Vol. 4, No. 12, pp. 89-96, 2015.

[10] M. K. K. Tripathi, "Analysis of Direct and Indirect Discrimination Discovery and Prevention Algorithms in Data Mining", *International Journal of Engineering Research and Technology*, Vol. 3, No. 2, pp. 2421-2425, 2014.

[11] H. M. Saindane and V. L. Kolhe, "Direct and Indirect Discrimination Prevention Approach using Association Rule Hiding", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 3, No. 6, pp. 55573-5581, 2015.

[12] A. Sahu, M. Dhakar, and P. Rani, "Comparative Analysis of Apriori Algorithm based on Association Rule", *International*

Journal of Computer Science and Communication, Vol. 6, No. 2, pp. 18-21, 2015.

[13] P. Gawade and J. Chaudhari, "A Comparative Survey of Discrimination Prevention Methods in Data Mining", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 5, No. 1, pp 1057-1060, 2015.

[14] H. Jafarzadeh, R. R. Torkashvand, C. Asgari, and A. Amiry, "Provide a new approach for mining fuzzy association rules using apriori algorithm" *Indian Journal of Science and Technology*, Vol. 8, No. 8, pp. 127-134, 2015.

[15] J. Manimaran and T. Velmurugan, "Analyzing the Quality of Association Rules by Computing an Interestingness Measures", *Indian Journal of Science and Technology*, Vol. 8, No. 15, pp. 1-12, 2015.

[16] S. Wachs, G. K. Jiskrova, A. T. Vazsonyi, K. D. Wolf, and M. Junger, "A cross-national study of direct and indirect effects of cyber bullying on cyber grooming victimization via self-esteem", *Psicologia Educativa*, Vol. 22, No.1, pp. 61-70, 2016.

[17] K. K. Tripathi, "Discrimination Prevention with Classification and Privacy Preservation in Data mining", P*rocedia Computer Science*, Vol. 79, pp.244-253, 2016.

[18] R.T. Dive1 and A.P. Khedkar, "An Approach for Discrimination Prevention in Data Mining", *International Journal of Application or Innovation in Engineering & Management*, Vol. 3, No. 6, 2014.

[19] M. S. Naveen and M. Shoerio, "Multilevel Anti-Discrimination and Privacy Preservation Correlativity", *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. 3, No. 7, pp. 4369-4376, 2015.

[20] P. Cheng, J. S. Pan, and C. W. L. Harbin, "Use EMO to protect sensitive knowledge in association rule mining by removing items", In: *Proc. of the 2014 IEEE Congress on Evolutionary Computation*, pp. 1108-1115, 2014.

[21] G. Sinha and S. M. Ghosh, "Identification of best algorithm in association rule mining based on performance", *International Journal of Computer Science and Mobile Computing*, Vol. 3, No. 11, pp. 38-45, 2014.

[22] L. Dong, R. Wang, and Y. Li. "Mining Association Rules Based on Certainty", *International Journal of Intelligent Engineering and Systems*, Vol. 5, No. 3, pp. 19-27, 2012.

[23] A. Dahbi, S. Jabri, Y. Ballouki, and T. Gadi, "A new method to select the interesting association rules with multiple criteria", *International Journal of Intelligent Engineering and Systems*, Vol. 10, No. 5, pp. 191-200, 2017.

[24] K. K. Tripathi and N. S. Chaudhari, "Mining Association rules and Differential Privacy Preservation using Randomization", *International Journal of Computer Sciences and Engineering*, Vol. 4, No.7, pp. 30-38, 2016.