IM_LR: An approach for Direct and Indirect Discrimination Prevention

Manoj Ashok Wakchaure, S.S. Sane

Abstract: ***Discrimination and privacy preservation are major challenges of data mining. Technique based on impact minimization to prevent discrimination has been reported in the literature. The technique computes fitness of generated frequent rules based on their antecedent, a pre-defined threshold and discrimination measure 'elift' to modify discriminating rules. This paper deals with a method called 'IMLR'. IMLR computes fitness of generated frequent rules based on their antecedent (attributes on left hand side of the rule) as well as consequences (class label on right hand side of the rule), a pre-defined threshold and offers selection of desired discrimination measures such as 'elift', 'slift', 'olift' etc. to modify discriminating rules. Experimentation results carried out using two well-known datasets 'Adult' and 'German' show that IMLR when used with certain discrimination measure provides better results in terms of various performance parameters such as DDPD, DDPP, IDPD, IDPP, Missed cost and Ghost cost when compared with reported technique.***

Index Terms: Data Quality, Direct and indirect discrimination, Discrimination measures.

I. INTRODUCTION

In social sense, discrimination refers to an action based on prejudice resulting in unfair treatment of people, where the distinction between people is operated on the basis of their membership to a category or minority, without regard to individual merit or circumstances. Classification without discrimination [12] is the approach for remove the discrimination. Examples of social discrimination include racial/ethnic, religious, gender, nationality, disability, and age-related discrimination; a large body of international laws and regulations prohibit discrimination in socially-sensitive decision making tasks, including credit scoring/approval, house lending, and personnel selection. In order to prove (or disprove) a discrimination charge before a court [21], or to perform a social analysis of discrimination in a given context, it is clearly needed to rely on quantitative measures of the phenomenon under study: for this reason, discrimination has been the subject of a large body of research in legal, economic and social sciences [19], as well as the subject of empirical analysis in a large number of juridical cases.

Discrimination can be either direct or indirect (also called systematic) [6] [7] [8]. Direct discrimination comprises of standards or strategies that unequivocally specify minority or burdened gatherings in light of delicate oppressive credits identified with aggregate participation. Indirect discrimination comprises of tenets or techniques that, while not unequivocally call attention to unfair viewpoints, intentionally or accidentally could produce prejudicial choices. Redlining by money related organizations (declining to allow home loans or protections in urban territories they consider as decaying) is an original case of indirect discrimination, albeit absolutely not alone. With a slight manhandle of dialect for conservativeness, in this paper indirect discrimination will likewise be alluded to as redlining and principles causing indirect discrimination will be called redlining arrangements. Indirect discrimination could happen since of the availability of some foundation information (rules) [14], for instance, that a specific postal district relates to a weakening region or a region with generally dark populace. The surroundings data may be accessible from freely accessible information (e.g., enumeration information) or may be gotten from the first informational collection itself as a result of the presence of nondiscriminatory traits that are profoundly associated with the delicate ones in the original data set. There is too much discrimination has seen in the real time data which is described in [16] [17].

II. RELATED WORK

In data mining, classification models are created on the base of old data for distinctive among components of dissimilar classes [9], in order to reveal the causes of class membership, or to forecast it for unsystematic models. In whichever cases, cataloging models can be accepted as a support to result making, as well as in generally sensitive work such as the admission of candidates to benefits, to public facilities, and to credit. It is apparent that relying on mined models for result making does not put ourselves on the secure side. Rather, learning from old data may mean to determine customary preconceptions that are harmful in realism, and to assign to such carry out the status of general rules, maybe intuitively, as these rules can be extremely unseen within a classifier. How the system are aware with discrimination has proposed in [18].

The problems of direct and indirect discernment anticipation in data mining by developing an attitude having a place with preprocessing stratagem for separation counteractive action [1]. They grew new systems material for immediate or backhanded separation anticipation exclusively or at the similar time, as an augmentation to their prior work [2]. The customary methodologies for separation anticipation consider just a single unfair trait. They can achieve either coordinate segregation or aberrant separation yet not both. This approach utilized different strategies to clean preparing informational indexes and outsourced informational indexes such that direct as well as aberrant biased choice tenets are transformed over to true blue (non-oppressive) characterization rules.

Revised Manuscript Received on May 06, 2019

Mr. Manoj A. Wakchaure, Research Scholar, Department of Computer Engineering, KKWIEER, Nashik, SPPU, Pune, India.

Dr. Shirish S.Sane, Professor & Head, Department of Computer Engineering, KKWIEER, Nashik, SPPU, Pune, India.

It incorporated the accompanying stages: Discrimination estimation and Data Transformation [13].

First statement the discernment problematic from the perspective of learning disclosure from databases which is describe in [3]. The question of segregation in information mining was tended to in a lead based way, by presenting the thought of biased characterization rules, with a specific end goal to recognize the probable risks of separation. This approach has a place with the preprocessing stratagem for segregation avoidance. It has been demonstrated that separation might be covered up in learning disclosure models removed from databases, and cataloging rule models are considered[22]. The advanceda new solution to the CND problematic by presenting a inspecting plans for influencing the information segregation to free as opposed to relabeling the dataset. This approach depended on preprocessing stratagem for separation counteractive action in [4]. At that point, in opinion of the sterilized information, a nondiscriminatory model could be scholarly. Since this model was found out on non-unfair information, it lessened the biased conduct for future characterization [14]. It helped in acquiring great outcomes with both steady and shaky classifiers, and alleviated the discrimination level by maintaining high accuracy level.

According [5] presents the creation of a decision tree classifier without segregation. Given a named dataset and a delicate property, the objective of separation mindful arrangement was to take in a classifier for anticipating the class name that did not segregate as for the touchy characteristic. Such limitations are called independency constraints. In [6] accessible a reformed Naive Bayes arrangement approach. Salvatore et al. [7] developed the DCUBE framework in dainty of a current approach of separation counteractive action.

III. PREVENTION OF DIRECT AND INDIRECT DISCRIMINATION

We present our approach, including the data transformation methods that can be used for direct and/or indirect discrimination prevention. For each method, its algorithm and its computational cost are specified. Direct and indirect discrimination prevention can be described.

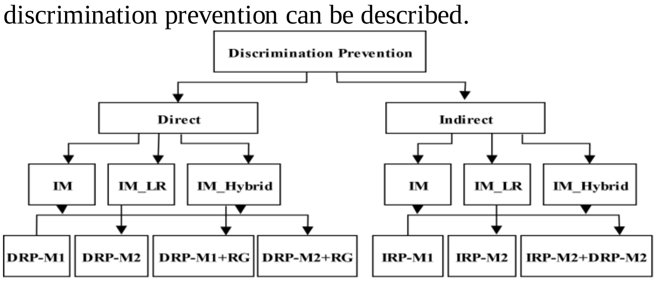

Fig. 1: Taxonomy of Discrimination prevention

The figure 1 shows the taxonomy of data discrimination of proposed system. This approach has been extended to encompass statistical significance of the extracted patterns of discrimination in and to reason about affirmative action and favoritism. Current discrimination discovery methods consider each rule individually for measuring discrimination without considering other rules or the relation between them.

IV. SYSTEM ARCHITECTURE

The given research work essence on discernment anticipation built on pre-processing, since the pre-processing method look like the utmost the one: it does not need varying the customary data mining algorithms, unlike the in treating method, and it agree data issuing (rather than just knowledge publishing), unlike the post dispensation method. The suggest work overawed the restraint based on pre-processing publish so far. In the suggest work new data alteration approaches are based on procedures for both direct and indirect discernment and can deal with numerous discriminatory items. This suggests method assurance that the converted data set is surely discernment allowed. It consists of measure to assess how much discernment has been removed and how much info loss has been sustained. Hence, the suggest work method to discernment anticipation is broader than in preceding work. Recommend work current a unified method to direct and indirect discernment prevention, with confirmed algorithms and all probable data conversion procedures that could be applied for direct or indirect discernment anticipation also identify the dissimilar structures of every process. The proposition techniques established new metrics that identify which data should be reformed, how many data should be different, and how those records should be different during data transformation. In adding, novel utility methods to calculate the dissimilar suggested discernment preclusion procedures in terms of data superiority and discernment elimination for together direct and indirect discernment. Based on the planned measures, present widespread investigational effects and compare the different possible methods for direct or indirect discernment anticipation to find out which procedures could be more successful in relations of low info loss and high discrimination removal.

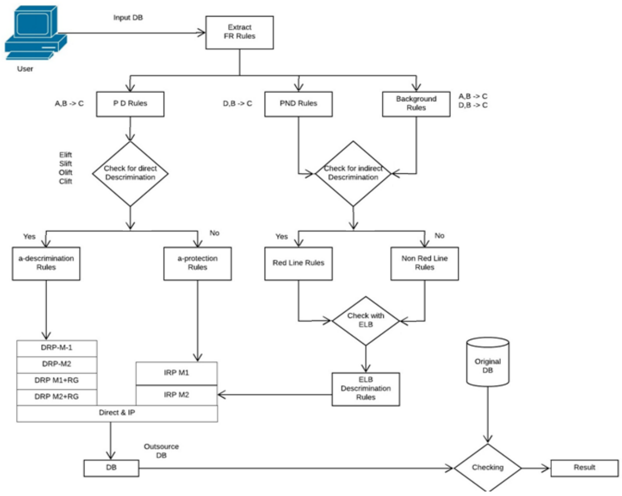

Fig. 2: System Architecture

V. EXPERIMENTAL ANALYSIS

Below tables shows the execution result with adult & German dataset. The Direct Rule Protection (DRP) and Indirect Rule Protection (IRP) has used for both approach. It used the fitness function from GA to calculate the weight.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The Table 1 to Table 7 shows the results using Elift,slift,olift and clift for Adult and Table 8 to Table 14 shows the results for German credit data respectively. The results for minimum support 2 percent and minimum confidence 10 percent for adult. It is improve the MC and GC rates than base approach [1]. The experiment has done with different alfa values on adult dataset using slift,elift,olift,clift measures.

Also the results for minimum support 5 percent and minimum confidence 10 percent for German credit. It is improve the MC and GC rates than base approach [1].The experiment has done with different alfa values on German Credit dataset using slift,elift,olift and clift measures.

All the Measures are taken from literature [1][19]and utility measures from[1]for our calculations.

A. Datasets

- Adult data set

The first experiment has used adult dataset which is taken from UCI Machine learning laboratory .data set having 48832 rows and for the train portion 32551 and for the test portion 16281 rows respectively. This dataset having 15 attributes with class attributes .class attributes is decision making attributes. Numerical attributes converted into categorical one so age above 30 and below or equal to 30.

- German credit data set

Basically it has used for second experiment .data set having 1000 rows and for the train portion 700 and for the test portion 300 rows respectively. This dataset having 21 attributes with class attributes .class attributes is decision making attributes. Numerical attributes converted into categorical one so age above 50 and below or equal to 50

Table I: Performance Comparison of DRP M-1 with Support of 2% and Confidence of 10% For Adult Dataset using IM_LR

Methods	Alfa	No. direct Alfa Discrimination Rules	Discrimination Removal		Data Quality		Execution Time in Sec.
			Direct				
			DDPD	DDPP	MC	GC	
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	66.08	0	NA
DRP M-1 as Reported in Literature	1.2	274	100	100	4.16	4.15	NA
DRP M-1 Using Elift	1.2	274	100	100	3.69	3.83	180
DRP M-1 Using slift	1.2	273	100	100	3.00	3.10	181
DRP M-1 Using clift	1.2	278	100	100	4.89	5.14	189
DRP M-1 Using Olift	1.2	275	100	100	3.12	3.22	189
No. of Frequent classification Rules: 5092				No. of background Rules: 2089			

Observations: when value of alpha is 1.2 then MR rules are 273, 278 & 275 using slift, clift and olift respectively. slift provides best values for MC and GC of 3 and 3.10 percent respectively and it is lesser than that of values obtained using IM. Execution time required with elift measure is minimum.

Table II: Performance Comparison of DRP M-2 with Support of 2% and Confidence of 10% For Adult Dataset using IM_LR

Methods	Alfa	No. direct Alfa Discrimination Rules	Discrimination Removal		Data Quality		Execution Time in Sec.
			Direct				
			DDPD	DDPP	MC	GC	
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	66.08	0	NA
DRP M-2 as Reported in Literature	1.2	274	100	100	0	0	NA
DRP M-2 Using Elift	1.2	274	100	100	0	0	182
DRP M-2 Using slift	1.2	273	100	100	0.14	0.14	183
DRP M-2 Using clift	1.2	278	100	100	0.39	0.39	185
DRP M-2 Using Olift	1.2	275	100	100	0.67	0.67	185
No. of Frequent classification Rules: 5092				No. of background Rules: 2089			

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Observations: Performance of both IM and IM_LR for both Discrimination removal and data quality with elift is identical and better when compared with IM_LR with other lifts.

Table III: Performance Comparison of DRP M-1+RG with Support of 2% and Confidence of 10% For Adult Dataset using IM_LR

Methods	Alfa	p	No. direct Alfa Discrimination Rules	Discrimination Removal		Data Quality		Execution Time in Sec.
				Direct				
				DDPD	DDPP	MC	GC	
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	NA	66.08	0	NA
DRP M-1+RG as Reported in Literature	1.2	0.9	274	100	100	4.1	4.1	NA
DRP M-1+RG Using Elift	1.2	0.9	274	100	100	4.1	4.28	204
DRP M-1+RG Using slift	1.2	0.9	273	100	100	3.75	3.90	214
DRP M-1+RG Using clift	1.2	0.9	278	100	100	4.75	4.99	222
DRP M-1+RG Using Olift	1.2	0.9	275	100	100	4.32	4.52	222
No. of Frequent classification Rules: 5092					No. of background Rules: 2089			

Observations: slift measure provides the best values for MC and GC of 3.75 and 3.90 respectively as compared to other lifts. However, needs slightly higher execution time when compared with elift.

Table IV: Performance Comparison of DRP M-2+RG with Support of 2% and Confidence of 10% For Adult Dataset using IM_LR

Methods	Alfa	p	No. direct Alfa Discrimination Rules	Discrimination Removal		Data Quality		Execution Time in Sec.
				Direct				
				DDPD	DDPP	MC	GC	
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	NA	66.08	0	NA
DRP M-2+RG as Reported in Literature	1.2	0.9	274	91.58	91.58	0	0	NA
DRP M-2+RG Using Elift	1.2	0.9	274	96.35	96.35	0	0	235
DRP M-2+RG Using slift	1.2	0.9	273	94.51	94.51	0	0	223
DRP M-2+RG Using clift	1.2	0.9	278	97.48	97.48	0.10	0.10	229
DRP M-2+RG Using Olift	1.2	0.9	275	96.00	100	0.22	0.22	229
No. of Frequent classification Rules: 5092					No. of background Rules: 2089			

Observations: IM_LR with Elift measure provides best values for all performance parameters compared to IM_LR with other lifts as well as IM with elift.

Table V: Performance Comparison of IRP M-1 with Support of 2% and Confidence of 10% For Adult Dataset using IM_LR

Methods	Alfa	No. Red Lining Rules	No. indirect Alfa Discrimination Rules	Discrimination Removal		Data Quality		Execution Time in Sec.
				Indirect				
				IDPD	IDPP	MC	GC	

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	NA	66.08	0	NA
IRP M-1 as Reported in Literature	1.1	21	30	100	100	0.54	0.38	NA
IRP M-1 Using Elift	1.1	21	30	100	100	0.71	0.71	265
IRP M-1 Using slift	1.1	21	30	100	100	0.77	0.77	255
IRP M-1 Using clift	1.1	21	30	100	100	0.59	0.59	267
IRP M-1 Using Olift	1.1	21	30	100	100	0	0	267
No. of Frequent classification Rules: 5092					**No. of background Rules: 2089**			

Observations: When we are considering alpha is 1.1 then olift gives maximum RR and Indirect alpha disc-rules .olift gives best result for data quality because no rules are lost in this particular measure.alsoslift measure required minimum execution time 255 seconds which is less as compared with others.

Table VI: Performance Comparison of IRP M-2 with Support of 2% and Confidence of 10% For Adult Dataset using IM_LR

Methods	Alfa	No. Red Lining Rules	No. indirect Alfa Discrimination Rules	Discrimination Removal		Data Quality		Execution Time in Sec.
				Indirect		MC	GC	
				IDPD	IDPP			
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	NA	66.08	0	NA
IRP M-2as Reported in Literature	1.1	21	30	100	100	0	0	NA
IRP M-2 Using Elift	1.1	21	30	100	100	0	0	240
IRP M-2 Using slift	1.1	21	30	100	100	0	0	243
IRP M-2 Using clift	1.1	21	30	100	100	0	0	251
IRP M-2 Using Olift	1.1	21	30	95.24	100	0	0	251
No. of Frequent classification Rules: 5092					**No. of background Rules: 2089**			

Observations: slift and clift provide better results when IDPD and IDPP are consider which is 100 Percent. eliftrequired less timei.e 240 seconds for execution so it is best for execution than all others.

Table VII: Performance Comparison of DRP M-2 +IRP M-2 with Support of 2% and Confidence of 10% For Adult Dataset using IM_LR

Methods	Alfa	No. Red Lining Rules	No. indirect Alfa Discrimination Rules	No. direct Alfa Discrimination Rules	Discrimination Removal				Data Quality		Execution Time in Sec.
					Direct		Indirect		MC	GC	
					DDPD	DDPP	IDPD	IDPP			
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	NA	NA	NA	NA	66.08	0	NA
DRP M-2+IRP M-2 as Reported in Literature	1.1	21	30	280	100	100	100	100	0	0	NA
DRP M-2+IRP M-2 Using Elift	1.1	21	30	280	100	100	100	100	0	0	305
DRP M-2+IRP M-2 Using slift	1.1	21	30	282	100	100	100	100	0.10	0	317

DRP M-2+IRP M-2 Using clift	1.1	21	30	283	99.64	100	100	100	0.16	0	319
DRP M-2+IRP M-2 Using Olift	1.1	21	30	282	100	100	100	100	0	0	319
No. of Frequent classification Rules: 5092					**No. of background Rules: 2089**						

Observations: Performance of IM_LR when used either with elift or olift is identical to that of performance of IM with elift. Use of elift measure needs lesser execution time as compared to the time needed in case of Olift, as more computations are needed in case of olif

Table VIII: Performance Comparison of DRP M-1 with Support of 5% and Confidence of 10% For German Credit Dataset using IM_LR

Methods	**Alfa**	**No. indirect Alfa Discrimination Rules**	**Discrimination Removal**		**Data Quality**		**Execution Time in Sec.**
			Direct				
			DDPD	**DDPP**	**MC**	**GC**	
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	64.35	0	NA
DRP M-1 as Reported in Literature	1.2	991	100	100	15.44	13.52	NA
DRP M-1 Using Elift	1.2	991	100	100	14.99	11.20	187
DRP M-1 Using slift	1.2	1003	100	100	15.41	13.66	198
DRP M-1 Using clift	1.2	1022	100	100	15.42	13.52	179
DRP M-1 Using Olift	1.2	1017	100	100	14.66	12.66	191
No. of Frequent classification Rules: 32340				**No. of background Rules: 22763**			

Observations: clift measure gives maximum MR rules as compared with other measures i.e 1022.olift and elift gives good results for MC and GC respectively because Data loss are 14.66 and 14.99 ,if MC is consider. If we are considering GC then 11.20 & 12.66 Percent of rules are lost in elift and olift respectively also clift takes 179 seconds execution time for execution which is very less comparatively.

Table IX: Performance Comparison of DRP M-2 with Support of 5% and Confidence of 10% For German Credit Dataset using IM_LR

Methods	**Alfa**	**No. direct Alfa Discrimination Rules**	**Discrimination Removal**		**Data Quality**		**Execution Time in Sec.**
			Direct				
			DDPD	**DDPP**	**MC**	**GC**	
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	64.35	0	NA
DRP M-2 as Reported in Literature	1.2	991	100	100	0	4.06	NA
DRP M-2 Using Elift	1.2	991	100	100	0	3.27	192
DRP M-2 Using slift	1.2	1003	100	100	0	3.77	194
DRP M-2 Using clift	1.2	1022	100	100	0	4.06	178
DRP M-2 Using Olift	1.2	1017	100	100	0	3.92	196
No. of Frequent classification Rules: 32340				**No. of background Rules: 22763**			

Observations: all measures are good for discrimination removal because it shows results are are 100 percent except DDPP in elift but for data quality elift measure is best because it preserve data quality 100 percent in MC and only 3.27 % lost in GC which is good than literature value and other measures also .clift required less time for execution ,it is 178 seconds so it is less than other measures.

Table X: Performance Comparison of DRP M-1+RG with Support of 5% and Confidence of 10% For German Credit Dataset using IM_LR

Methods	Alfa	P	No. indirect Alfa Discrimination Rules	Discrimination Removal		Data Quality		Execution Time in Sec.
				Direct		MC	GC	
				DDPD	DDPP			
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	NA	64.35	0	NA
DRP M-1+RGas Reported in Literature	1.2	0.9	991	100	100	13.34	12.01	NA
DRP M-1+RG Using Elift	1.2	0.9	991	100	100	11.54	9.82	201
DRP M-1+RG Using slift	1.2	0.9	1003	100	100	14.02	11.22	207
DRP M-1+RG Using clift	1.2	0.9	1022	99.12	99.10	13.33	12.10	201
DRP M-1+RG Using Olift	1.2	0.9	1017	100	100	13.02	11.58	203
No. of Frequent classification Rules: 32340				No. of background Rules: 22763				

Observations: elift,slift and olift gives best results for discrimination Removal in Direct methods .elift measure is good for data preservation and execution time required than others.elift gives better results than literature which are 11.54 and 9.82 respectively in MC and GC .using elift and clift measures ,less execution time required than other i.e 201 seconds.

Table XI: Performance Comparison of DRP M-2+RG with Support of 5% and Confidence of 10% For German Credit Dataset using IM_LR

Methods	Alfa	P	No. direct Alfa Discrimination Rules	Discrimination Removal		Data Quality		Execution Time in Sec.
				Indirect		MC	GC	
				IDPD	IDPP			
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	NA	64.35	0	NA
DRP M-2+RGas Reported in Literature	1.2	0.9	991	100	100	0.01	4.06	NA
DRP M-2+RG Using Elift	1.2	0.9	991	99.80	100	0	4.126	246
DRP M-2+RG Using slift	1.2	0.9	1003	100	100	0.01	4.01	231
DRP M-2+RG Using clift	1.2	0.9	1022	100	100	0	4.06	219
DRP M-2+RG Using Olift	1.2	0.9	1017	100	100	0.03	4.05	246
No. of Frequent classification Rules: 32340				No. of background Rules: 22763				

Observations: for MC,elift and clift are best because it protect 100% rules so it shows o loss but for GC ,slift is good than others because it gives 4.01 which is less than any other and literature also.clift measure needs 219 seconds time for execution of algorithm which is less .olift measure is also giving good result than literature because it is 4.05.

Table XII: Performance Comparison of IRP M-1 with Support of 5% and Confidence of 10% For German Credit Dataset using IM_LR

Methods	Alfa	No. Red Lining Rules	No. indirect Alfa Discrimination Rules	Discrimination Removal		Data Quality		Execution Time in Sec.
				Indirect		MC	GC	
				IDPD	IDPP			
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	NA	64.35	0	NA
IRP M-1 as Reported in Literature	1	37	42	100	100	1.62	1.47	NA

IRP M-1 Using Elift	1	37	42	100	100	1.29	1.31	264
IRP M-1 Using slift	1	37	42	100	100	1.77	1.80	289
IRP M-1 Using clift	1	37	42	100	100	1.53	1.55	271
IRP M-1 Using Olift	1	37	42	100	100	1.44	1.46	271
No. of Frequent classification Rules: 32340					**No. of background Rules: 22763**			

Observations: elift is good for MC because only 1.29 percent of rules has been loosed. The value in literature is 1.62 and elift is good for GC because it gives value 1.31 which is low than others in GC. Also less execution time required for elift measure i.e 264 seconds.

Table XIII: Performance Comparison of IRP M-2 with Support of 5% and Confidence of 10% For German Credit Dataset using IM_LR

Methods	Alfa	No. Red Lining Rules	No. indirect Alfa Discrimination Rules	Discrimination Removal		Data Quality		Execution Time in Sec.
				Indirect		MC	GC	
				IDPD	IDPP			
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	NA	64.35	0	NA
IRP M-2as Reported in Literature	1	37	42	100	100	0	0.96	NA
IRP M-2 Using Elift	1	37	42	100	100	0	0.96	259
IRP M-2 Using slift	1	37	42	100	100	0	0.36	274
IRP M-2 Using clift	1	37	42	100	100	0	0.95	260
IRP M-2 Using Olift	1	37	42	100	100	0	1.03	262
No. of Frequent classification Rules: 32340					**No. of background Rules: 22763**			

Observations: all measures are giving best results for indirect discrimination removal because it showing value 100 with respect to various measures but slift provide best result for MC and GC because MC and GC values are 0 and 0.36 respectively .elift needs less execution time i.e 259 seconds which is less than others

Table XIV: Performance Comparison of DRP M-2 +IRP M-2 with Support of 5% and Confidence of 10% For German Credit Dataset using IM_LR

Methods	Alfa	No. Red Lining Rules	No. indirect Alfa Discrimination Rules	No. direct Alfa Discrimination Rules	Discrimination Removal				Data Quality		Execution Time in Sec.
					Direct		Indirect		MC	GC	
					DDPD	DDPP	IDPD	IDPP			
Removing Disc Attributes as reported in Literature	NA	NA	NA	NA	NA	NA	NA	NA	64.35	0	NA
DRP M-2+IRP M-2 as Reported in Literature	1	37	42	499	99.97	100	100	100	0	2.07	NA
DRP M-2+IRP M-2 Using Elift	1	37	42	499	99.90	100	100	100	0	2.26	333
DRP M-2+IRP M-2 Using slift	1	37	42	503	99.90	100	100	100	0	1.86	341
DRP M-2+IRP M-2 Using clift	1	37	42	497	100	99.75	100	100	0	2.00	325
DRP M-2+IRP M-2 Using Olift	1	37	42	500	100	100	100	100	0	2.23	325
No. of Frequent classification Rules: 32340					**No. of background Rules: 22763**						

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

Observations: Number of RR rules are same for all measures i.e 37.but elift gives more indirect rules and slift gives more MR rules 42 and 503 respectively. elift and slift gives same results for DDPD and others which is 99.90 and 100 in DDPD & DDPP respectively. Slift is best for Data quality Preservation which is 0 and 1.86 in MC and GC respectively. Clift and olift required same execution time for execution which is less than others 325 seconds.

VI. CONCLUSION AND FUTURE SCOPE

Experimentation was one on IM_LR Method for both Popular datasets Adult and German Credit. Modified Algorithms gives better results but not single measure independently Perform better. Mostly slift&elift gives best values for our Parameters. In future we will working on IM_Hybrid method for testing the results and comparison . Future work will include the exploration of relationship between expectation of discrimination and conservation of privacy in data mining with different dataset and write a new weight calculation algorithm like ACO, NN etc for getting best accuracy.

REFERENCES

1. Sara Hajian and Josep Domingo-Ferrer A Methodology for Direct and Indirect Discrimination Prevention in Data Mining, Data Mining and Knowledge Discovery, vol. 25, no. 7, pp. 1445-1459, 2013
2. R. Agrawal and R. Srikant, Fast Algorithms for Mining Association Rules in Large Databases, Proc. 20th Intl Conf. Very Large Data Bases, pp. 487-499, 1994.
3. V. Verykios and A. Gkoulalas- Divanis, A Survey of Association Rule Hiding Methods for Privacy, PrivacyPreserving Data Mining, Models and Algorithms, C.C.Aggarwal and P.S. Yu, Springer, 2008.
4. D. Pedreschi, S. Ruggieri, and F. Turini, Measuring Discrimination in Socially-Sensitive Decision Records, Proc. Ninth SIAM Data Mining Conf. (SDM 09), pp.581-592, 2009.
5. F. Kamiran and T. Calders, Classification without Discrimination, Proc. IEEE Second Intl Conf. Computer, Control and Comm.(IC4 09), 2009.
6. 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.
7. F. Kamiran and T. Calders, Classification with no Discrimination, by Preferential Sampling, Proc. 19th Machine Learning Conf. Belgium and The Netherlands,2010.
8. D. Pedreschi, S. Ruggieri, and F. Turini, "DiscriminationAware Data Mining," Proc. 14th ACM Int'l Conf. Knowledge Discovery and Data Mining (KDD '08), pp. 560-568, 2008.
9. F. Kamiran, A. Karim, a nd X. Zhang. Decision theory for discrimination-aware classication.In In ICDM 2012, pp. 924929. IEEE Computer Society, 2012.
10. Hajian, S., Domingo-Ferrer, J. & Martinez-Ballesté, A. (2011b). Rule protection for indirect discrimination prevention in data mining. Modeling Decisions for Artificial Intelligence-MDAI 2011, Lecture Notes in Computer Science 6820, pp. 211-222.
11. Hajian, S., Domingo-Ferrer, J. & Martinez-Ballesté, A. (2011a).Discrimination prevention in data mining for intrusion and crime detection. Proc. of the IEEE Symposium on Computational Intelligence in Cyber Security (CICS 2011), pp. 47-54. IEEE.
12. Pedreschi, D., Ruggieri, S. ,Turini, F. , Integrating induction and deduction for finding evidence of discrimination. , Proc. of the 12th ACM International Conference on Artificial Intelli-gence and Law (ICAIL 2009), pp. 157-166. ACM.,2009b.
13. Manoj Ashok Wakchaure ,Prof.Dr.ShirishSane,"Discrimination Prevention by Different Measures in Direct Rule Protection Algorithm",IJETAE Exploring Research and Innovation 2015.
14. Kamiran, F., Calders, T. &Pechenizkiy, M. (2010). Discrimination aware decision tree learning. Proc. of the IEEE International Conference on Data Mining (ICDM 2010), pp. 869-874. ICDM.
15. Newman, D. J., Hettich, S., Blake, S. L. &Merz, C.J. (1998). UCI Repository of Machine Learning Databases. http://archive.ics.uci.edu/ml.
16. Parliament of the United Kingdom. (1975). Sex Discrimination Act.http://www.opsi.gov.uk/acts/acts1975/ PDF/ukpga 19750065 en.pdf.
17. Parliament of the United Kingdom. (1976). Race Relations Act. http://www.statutelaw.gov.uk/content.aspx? activeTextDocId=2059995.
18. B. Gao and B. Berendt. Visual data mining for higher-level patterns: Discriminationaware data mining and beyond. In Proc. of Belgian Dutch Conf. on Machine Learning (Benelearn 2011), pp. 4552, 2011.
19. Mr.Manoj Ashok Wakchaure,Prof.Dr.S.S.Sane(2018),"An Empirical Evaluation of Various Discrimination Measures for Discrimination Prevention",International Journal of Engineering &Techonology
20. Priya P, Pamila JM. Identification of Direct and Indirect Discrimination in Data Mining..
21. Ruggieri, S., Pedreschi, D. &Turini, F. (2010).Data mining for discrimination discovery. ACM Transactions on Knowledge Discovery from Data, 4(2) Article 9.
22. Manoj Ashok Wakchaure , and Prof.Dr.Shirish S. Sane. "A Novel Approach for Discrimination Prevention and Privacy Preservation in Data Mining." International Journal Of Advance Research And Innovative Ideas In Education 2.3(2016) : 3822-3827.

AUTHORS PROFILE

Mr. ManojA.Wakchaure, Received Bachelors Degree in Information Technology and Masters Degree M.Tech in Computer Engineering from Pune University,Pune&BhartiVidyapeeths Deemed University, Pune respectively. Currently he is pursuing his research work at research centre at K KWagh Institute of Engineering Education and Research, Nashik affiliated to the Savitribai Phule Pune University, Pune. His areas of interests include Data Mining, databases, Computer Network & Security, Information system.

Dr. ShirishS.Sane, obtained his Bachelors' Degree in Computer Engineering from the Pune Institute of Computer Technology (PICT), Pune, Masters Degree M. Tech (CSE) from IIT Bombay and Ph. D. in Computer Engineering from SavitribatiPhule Pune University, formerly known as University of Pune. Dr. Shirish is working as the Head of the Computer Engineering Department and Vice Principal at K KWagh Institute of Engineering Education & Research, Nashik. He has published more than 60 research papers at the National and International Conferences and Journals. He has also authored books on the subjects "Data Structures" and "Theory of Computer Science". His areas of interests include Data Mining, Databases, Compilers and Cloud Computing.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*