

DATA DISCRIMINATION PREVENTION IN CUSTOMER RELATIONSHIP MANAGMENT

Pradnya Deshpande¹, Rashi², Sneha Zanzane³

¹*Bhivarabai Sawant Institute of Technology and Research, Pune University of Pune*

²*Bhivarabai Sawant Institute of Technology and Research, Pune University of Pune*

³*Bhivarabai Sawant Institute of Technology and Research, Pune University of Pune*

Abstract

Data mining is a very important technology to derive important data occult in pompous heap of data. However, there is negative social awareness about data mining. It consists of unfairly treating people on the basis of their membership to a specific group. Automated data aggregation and data mining methods is useful in making automated decision such as customer identification and customer development. If partiality in training data sets is such a that regards sensitive attributes like gender, religion, color etc. Then discriminatory issues can arise. To remove this, antidiscrimination methods like discrimination discovery and prevention have been included in data mining. Direct and indirect are the two types of discrimination. When decisions are made based on discriminatory entities then, that discrimination is direct discrimination. When decisions are made based on non-discriminatory entities then that discrimination is indirect discrimination. Indirect discrimination corresponds to biased sensitive issues. Here we proposed discrimination prevention in data mining for customer relationship management. We proposed an enabling system supporting a business strategy for long term, profitable relationships with customers. Our paper is related to one-to-one marketing and loyalty programs. Our paper is related to the discrimination in online shopping system. Location based discrimination and dynamic pricing are the main objectives of our papers.

Keywords: Antidiscrimination, direct and indirect discrimination prevention, data mining, dynamic pricing, Apriori algorithm.

-----***-----

1. INTRODUCTION

Data mining is the process of extracting information from a set of data and transforming it into a suitable structure for future use. It includes raw analysis step, database and data management aspects. It is largely applied to any pattern of huge data or information processing. Multiple groups in data which can be used to obtain more accurate prediction results by decision support system are identified by data mining step .Data mining is very helpful in extracting important data from a large collection of database.

There are six common types of task in data mining [4].

Anomaly detection – anomaly detection is the verification of items, events that need further investigation. It is applicable to fraud detection

Association rule learning –it is used for finding relations between fluctuating variables in huge data sets.

Clustering –the grouping of identical objects is called clustering. It is used in sequence analysis.

Classification –process of identifying categories of new objects. It is used in drug discovery and development.

Regression – it is used to find methods of data with minimum errors.

Summarization – it gives summary of set of data which include generation of report.

Discrimination is nothing but unequal treatment that causes harm. Discrimination is of two types: direct and indirect. Direct discrimination can spotted easily .Direct discrimination happens when merit or ability are not the factors for the decision making. Instead of it is based on gender, race, religion or age. Indirect discrimination occurs when an employer impose certain requirements, practices and conditions that has an unfavorable impact on one group or other. With direct discrimination employee can argue that there may be discrimination but that is required for the job.

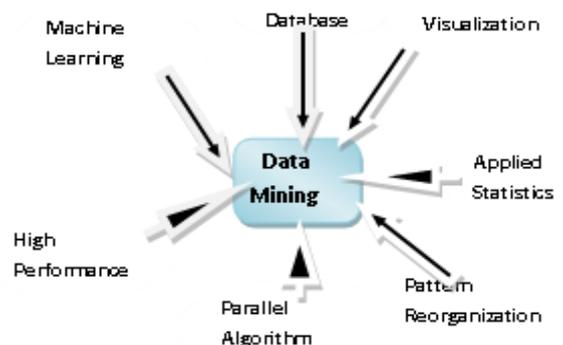


Fig.1. Data mining technique

Customer relationship management (CRM) is a technology for management of interaction between company and company’s current and future customers. It provides automation in the process. Operational and analytical are the two types of CRM framework. In operational CRM the automation of business process is done and in analytical CRM diagnosis of characteristic and behavior of customer is done. Analytical CRM is useful in better discrimination for an organization to allocate resources to group of customers which might be profitable for organization in future. In this paper we are presenting a methodology for preventing discrimination in customer relationship management for online shopping system. In online shopping website a large collection of data is present. The data contains information about various products of different categories. We cannot analyze them manually; this will cause wastage of time and information. Thus we can use the concept of data mining in this area. In online shopping system there may be possibilities of discrimination among the customers depending on the location.

2. METHODOLOGY

The Predreschi et al. was the first to propose the discovery of discriminatory decisions [5] [8]. The inductive part i.e. mining classification and the deductive part i.e. reasoning on them on quantitative measures of discriminations that forms the legal definition of discrimination are the major approaches of it. The antidiscrimination i.e. discrimination prevention in data mining contains the inclusion of patterns which will not cause any discriminatory decisions even though the there is a biased data sets are present.

Three methods are used for discrimination prevention:

2.1 Preprocessing:

Preprocessing is process of transforming source data. It transforms the data in such a way that discriminatory attributes contained in a source data are removed so that illegal decision rules are removed from the transformed data. After that we can apply any standard algorithm for data mining. The privacy preservation documentation can adapt the hierarchy based generalization and data transformation which are the approaches of the preprocessing. For this [6] , [7] performs the disciplined exaggeration of the training data sets from which classifier can make minimum modifications in the data sets to get fair data. Preprocessing is mostly used when data mining is to be performed by external parties (other than data holders). In the case of online shopping system when the customer places an order then the process of customer authentication is preprocess.

2.2 In-processing:

The data mining algorithms have to be changed in such a way that there should not be any inequitable rules of decision. For instance, to remove the discrimination from the source data set different technique is given in [9] where nondiscriminatory information is enclosed within the decision trees by providing a different approach for the

criteria for splitting and trimming strategy through approach of relabeling the novel leaf[1]. However, in-processing discrimination method must be built on some new unusual data mining algorithms. Here we cannot use standard data mining algorithms. There can be possibility of discrimination in in-processing in the availability of the product according to pin code of the customer.

2.3 Post-processing:

Instead of removing data from source data set or modifying the data mining algorithms, it is more suitable to modify the resulting data mining models. For instance, in [15], for classification rules confidence-altering methods are firmly proposed. Publishing of data sets are not allowed in the postprocessing technique. It only allows publishing the modified data mining models (judgmental publishing). Because of this only data holders can perform data mining. In the order processing, sometimes it may happen that if customer places the order and his/her payment is done the seller can take much time to deliver the product rather than cash on delivery.

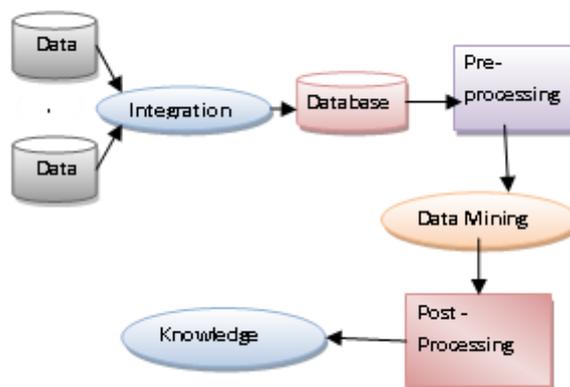


Fig.2. Knowledge discovery process

Preprocessing approach is the easiest technique for removing discrimination but it mostly causes important information loss from the source data set. Preprocessing is the technique which solves the problem of direct discrimination. But it causes indirect discrimination in most of the cases. There are some instances like pin code is not acceptable for some websites for delivery of the product, and there can be some instances that are exceptionally associated with the sensitive attributes (e.g. Race) which are responsible for the direct and indirect discrimination.

There are two crucial challenges related to discrimination prevention:

1. The first and important challenge is we should consider both direct and indirect discrimination and not only the direct discrimination.
2. The next challenge is to search for better settlement between removal of discrimination and quality of the transformed data set and data mining models.

Even though there are many already proposed techniques which are used for discrimination prevention for above stated methods (preprocessing, in-processing, post processing), discrimination prevention in customer relationship management for online shopping system stays a largely unfamiliar research approach. In this paper we have concentrated on discrimination prevention techniques which are very much useful for customers as well as organizations so that no one should face the discrimination. Here we have proposed our system for the discrimination which is relegated to the online shopping system. After analyzing and interpreting the data sets, the resulting models and correlations may become useful observation and tools for decision.

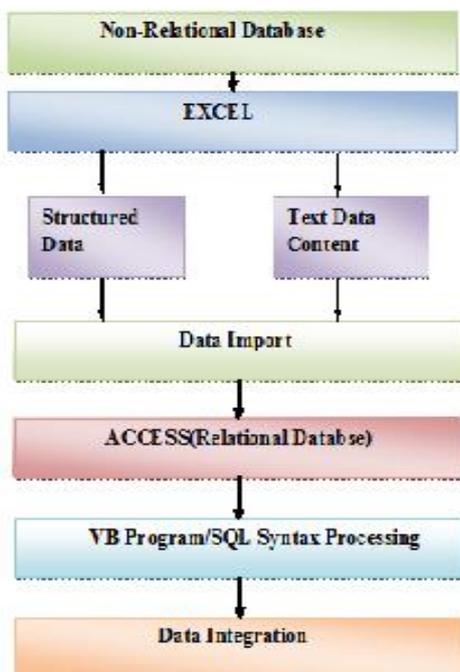


Fig. 3: Data processing

3. PLAN OF THIS PAPER

Here we are trying to discover and remove direct discrimination as well as indirect discrimination. In this paper we have discussed some methods for preprocessing for preventing discrimination in online shopping system. In online shopping system there can be possibility that there is unavailability of product which is requested by the customer at particular pin code provided. We have to check the availability of the product. Here indirect discrimination can happen if the website provides that this particular product is not available for the pin code. This can be prevented by using APRIORI algorithm which is already explained in some paper on discrimination prevention. By using this algorithm we avoid discrimination by providing some other sellers those can deliver the product at that pin code. This will prevent location based indirect discrimination. In some cases it can happen that product requested by customer is not available with the same specifications then instead of directly cancelling the order of the customer we can suggest

them the alternatives with either same specification but different seller or with same seller but with some other specifications.

Apriori algorithm formula applied to check reach-ability of seller

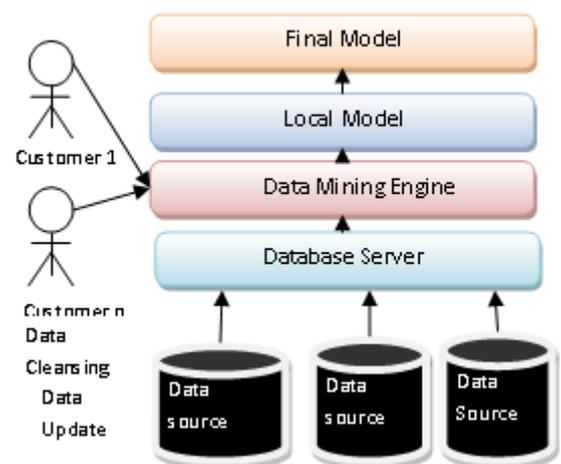
$$Conf (pincode \rightarrow seller) = \frac{supp(pincode, seller)}{supp(pincode)}$$

To check product availability

$$Conf (product \rightarrow pincode) = \frac{supp(product, pincode)}{supp(product)}$$

Discrimination based on user’s reliability is also done. Peoples who can’t receive the product for some reason are blacklisted. Customer suffers from direct discrimination. Price discrimination is common in online shopping. New users get more benefit than the regular ones. In this case dynamic programming is done. Some big name retailers are experimenting with offering different prices and product to different users. The websites like Amazon is using dynamic pricing for price discrimination. If we change our IP address then the price of product differs. If we delete the cache memory then also price differs.

When customer registers for the online shopping for the first time then it may happen that customers can provide same data in different ways. For example, pin code provided by the customer can be of two types: the first is pune-14 and second is 411014. Customer can enter any of this kind of pin code as both refer to the same area of the city Pune. But we are not provided the standard way for entering pin code of any one of the kind then there can be time consuming process required which will match the city name with the particular pin code for every time that customer requires the access.



standard form of database at the time of registration only. The standardization of the database gives us efficient data with less time consumption.

4. CONCLUSION AND FUTURE WORK

Peoples do not like to be discriminated because of their location, previous record and price etc., especially when these attributes are used for making decisions about providing them the order. The purpose of this paper is to identify different discrimination occurs in the online shopping system and how we can prevent those direct or indirect discrimination .To achieve this first identify categories where direct or indirect discrimination occurs in decision making process. And second step is to transform data to prevent discrimination .Finally discrimination free online shopping system can be produced without damaging data quality.

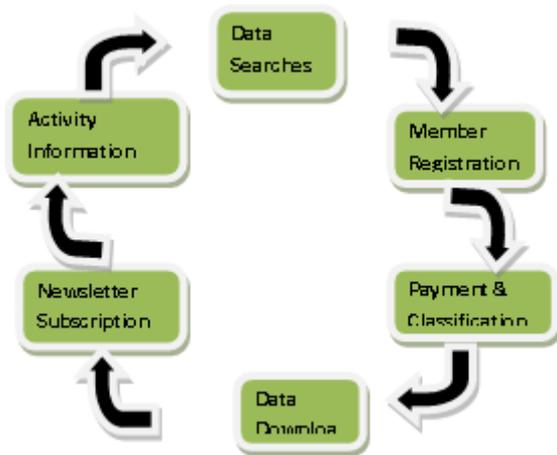


Fig.5. Operating model

In future we can work on predicting user behavior, analyze and give suggestions regarding each user and check the reliability of customers.

REFERENCES

[1] Sara Hajian, Josep Domingo-Ferrer, “A Methodology for Direct and Indirect Discrimination Prevention in Data Mining”, VOL. 25, NO. 7, JULY 2013

[2] E.W.T. Ngai, Li Xiu, D.C.K. Chau, “ Application of data mining techniques in customer relationship management : Aliterature review and clasiification”, Expert System with Application 36 (2009) 2592-2602

[3] Che-Wei Chang, Chin-Tsai Lin, Lian-Qing Wang, “Mining the text information to optimizing the customer relationship management“, Expert System with Application 36 (2009) 1433-1443

[4] Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases". Retrieved 17 December 2008

[5] D. Pedreschi, S. Ruggieri, and F. Turini, “Discrimination-Aware Data Mining,” Proc. 14th ACM Int’l Conf. Knowledge Discovery and Data Mining”, (KDD ’08), pp. 560-568, 2008

[6] .F. Kamiran and T. Calders, “Classification without Discrimination,” Proc. IEEE Second Int’l Conf. Computer, Control and Comm. (IC4 ’09), 2009

[7] .F. Kamiran and T. Calders, “Classification with no Discrimination by Preferential Sampling,” Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.

[8] S. Ruggieri, D. Pedreschi, and F. Turini, “Data Mining for Discrimination Discovery,” ACM Trans. Knowledge Discovery from Data, vol. 4, no. 2, article 9, 2010.

[9] 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.

[10] 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.