



Data Mining and its Various Concepts

¹Alisha Sikri, ²Surjeet Dalal, ³N.P Singh, ⁴Neeraj Dahiya

¹alisha.sikri92@gmail.com, ²profsurjeetdalal@gmail.com, ³knpsingh@mdi.ac.in

Abstract

Information mining is known as the extraction of concealed prescient data from expansive databases whose primary concentration is to enable organizations to concentrate on the most critical data to display in their information stockrooms. Data mining can in like manner be called as the examination of data and the use of the distinctive programming techniques for finding illustrations and regularities in the given courses of action of data. The articulation "Electronic exchange" (or web business) ordinarily implies the usage of an electronic medium to finish all the business trades. Numerous a times it alludes to the offer of items by means of Internet, however it likewise incorporates the acquiring of systems through Internet. This primary concentration of this paper is to give the fundamental presentation about the different information mining methods accessible and furthermore to break down these procedures based on their execution. The paper additionally characterizes the different goals of the information mining in internet business. The paper additionally concentrates on the bunching methods and tries to think about the different grouping procedures.

Keywords: - Data Mining, Web Mining, E-commerce, Clustering

1 Introduction

With the fast increment of the data accessible online alongside the developing prominence of electronic trade, web information mining is being given careful consideration. So as to consequently discover concentrate and channel data from the assortment of web assets and areas, we require another innovation and this innovation is web mining. Web mining is the use of information mining systems to find designs from the web. It consolidates together the different customary information mining innovation and web mining innovation together and can assume a critical part from various perspectives, similar to the mining of web crawlers, the advancement of web indexes, enhancing and upgrading the quality and productivity of web indexes etc¹.

E-Commerce² is a sort of plan of activity, or area of a greater arrangement of activity, that engages a firm. Affiliation or individual to lead business over an electronic framework, frequently the E-exchange is creating and winning omnipresence around the world in view of its distinctive purposes of intrigue like negligible exertion, it's favorable nature and besides it's brisk, safe, and strong trade and moreover it is free from prerequisite of time and space. Information mining in internet business is

the web mining which is utilized to consequently find and concentrate fascinating and helpful examples and certain data from www assets (web records) and conduct (Web administrations).

Data mining³ is the extraction of covered perceptive information from sweeping databases; it is a competent advancement which has a mind boggling potential to empower relationship to focus on the most basic information in their data appropriation focuses. The Data mining device anticipates the future examples and rehearses and moreover helps the relationship to settle on proactive data driven decisions. Information mining⁴ is advanced by the designer, clients, and scientists of awesome intrigue, and seems uncommon research information mining and the learning disclosure.

2 Data Mining

Information mining⁵ is a procedure to remove the certain, not known ahead of time and possibly helpful data and learning from an extensive number of inadequate, boisterous, dubious and arbitrary useful application information. The utilization of information mining strategies have been found in different fields like Business and Management, Education, back, restorative and wellbeing and social networks¹⁶. Data mining is a reliance on the application, various data mining applications may require unmistakable data mining methodologies, and the dealing with stream may similarly extraordinary, the general data mining process as showed up in fig1.

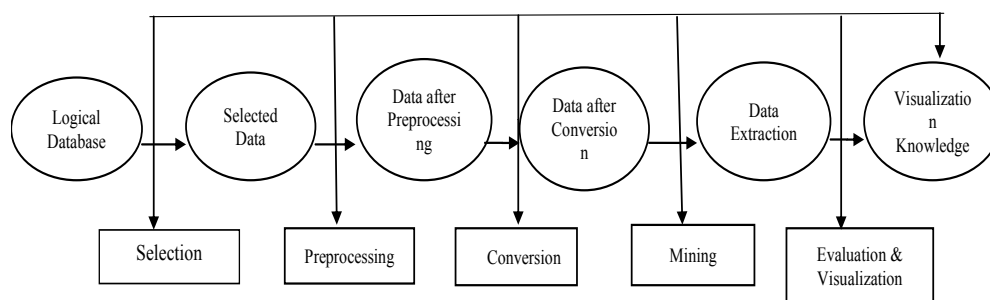


Fig. 1 Basic Process of Data Mining⁵

The information mining undertakings are of various sorts relying upon the utilization of information mining comes about the information mining assignments are characterized as³:-

1. **Exploratory Data Analysis:** It is basically investigation of the information with no unmistakable thoughts in context of we are searching for. These strategies are intelligent and good at illustration.
2. **Descriptive Modeling:** It portrays every one of the information. It joins models for general probability course of the data, allocating the p-dimensional space into social events and models delineating the associations between the components.
3. **Predictive Modeling:** This model allows the estimation of one variable to be anticipated from the known estimations of different factors.
4. **Discovering patterns and rules:** It worry with design discovery, the point is spotting deceitful conduct by recognizing locales of the space characterizing the distinctive sorts of exchanges where the information focuses essentially not the same as the rest.

.The various Data Mining algorithms are²:-

S.NO	Name of Algorithm
1.	Descriptive
2.	Association
3.	Clustering
4.	Estimation
5.	Predication

Table1: Data Mining Algorithms²

3 Web Data Mining

Web content mining depends on the site page content and is the procedure of data finding and learning removing from extensive measure of web information. Web content mining is partitioned into content mining (counting txt, HTML organization) and interactive media mining (counting pictures, sound and other media frames). As indicated by the distinctive mining focuses on the web mining can be isolated into following 3 categories²:-

- **Web content mining** mines or concentrates valuable data or information from Web page substance. For Example, we can naturally order and group Web pages as indicated by their themes. Be that as it may, we can likewise find designs in site pages to remove valuable information, for example, portrayals of items, postings of gatherings, and so on, for some reasons.
- **Web structure mining** finds valuable learning from hyperlinks (or connections for short), which speak to the structure of the Web. For instance, from the connections, we can find imperative Web pages, which, by chance, is a key innovation utilized as a part of web crawlers.
- **Web usage mining** otherwise called Web log mining, its primary goal is to discover fascinating model from the Web visit record. What's more, for the exploration here there are

two primary bearings: general access designs track and customized utilize record track.

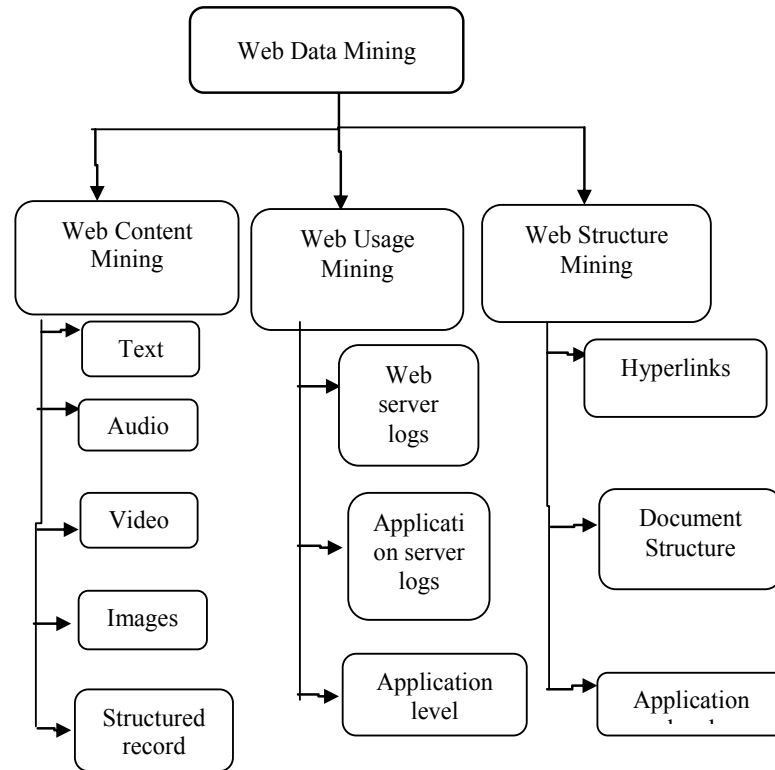


Fig.2 Categorization of Web Mining^{2,10}

There are various issues which are involved while applying Web mining to E-commerce. These are as follows¹⁴:-

1. There are sure critical parameters to be executed in retail an E-trade locale which incorporates the programmed time out of the client's session because of the apparent dormancy at the client end and it should be construct not absolutely in light of information mining calculations, but rather on the relative significance of the clients to the association.
2. Level of abstraction is vital.
3. The right level of granularity is imperative for data mining.
4. It is an excessive exercise to deliver the logs for a couple of million trades in this manner it is an adroit decision to make logs by driving self-assertive inspecting, as is done in quantifiable quality control. Thus, it is basic to make frameworks which use analyzing techniques in a keen outline.
5. The design of the user interface forms should consider the issues of data mining in mind.

4 Clustering

Grouping is characterized as the characterizations of a data-set into facilitate subsets (Clustering), it is done as such that the information in every subset shares some basic attribute frequently Proximity as per some characterized remove measure. Clustering is done to enhance the nature of the sites by gathering Comparative sites in groups⁸Users are known to have difficulties in dealing with information retrieval search outputs especially if the outputs are above a certain size. Clustering can enable them to find the relevant documents more easily and also help them to form an understanding of the different facets of the query that have been provided for their inspection. Clustering analysis consists of the following four step process:

- a. Data pretreatment,
- b. Define a distance function for measuring the Similarity between data
- c. clustering or group
- d. Appraisal output

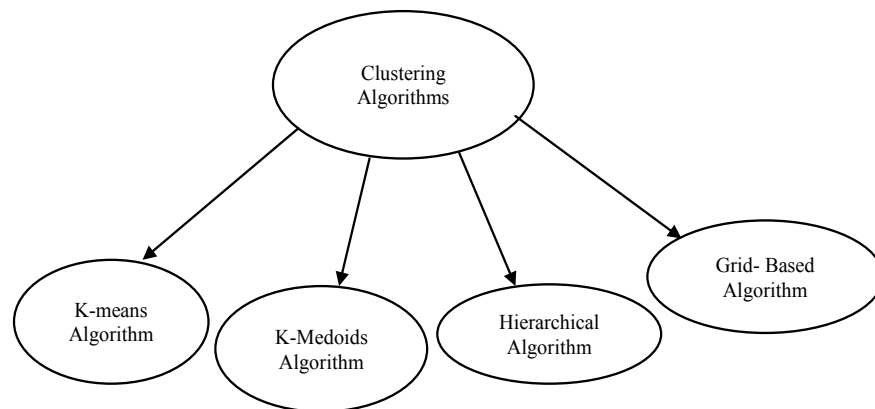


Fig3.Categorization of ClusteringAlgorithms⁹

S.NO	Name of Algorithm	Description	Time and Space Complexity	Benefits	Limitations
1.	K-means Algorithm	It is a way to organize the data items based on some features into k groups where k is a positive	Time Complexity:- $O(I*k*m*n)$ Space Complexity:- $O(m+k)n$	1. Clusters have similar density.	1. Sensitive to outliers 2. Deciding k

		integer.			
2.	K-Medoids Algorithm	In this medoids is calculated instead of mean.	Time Complexity:- $O(k(n-k)/2)$	1. It is stronger than K-means.	1. It is not good for noisy data.
3.	Distributed K-means Algorithm	This Clustering method uses normalized data.	Time Complexity:- $O(nkdi)$	1. The performance for normalized data is very good.	
4.	Hierarchical Algorithm	It is a process of cluster analysis which forms a hierarchy of clusters.	Time Complexity:- $O(n^2)$	1. It is very accurate.	1. Its speed is too slow for big data sets.
5.	Grid- Based Algorithm	This Algorithm makes grid.	Time Complexity:- $O(n^2)$	1. Fast-processing, Query-free. 2. Order is $O(k)$, where k is total grid cells at the lowest stage.	1. Size is limited to union of cells.

Table2: Clustering Algorithms and their Comparisons^{2, 10}

5 Literature review

[2]. Zhiwu Liu, Li Wang, “Study of Data Mining Technology Used for E-commerce”, the creator of this paper chiefly concentrates on the idea and arrangement of Web Mining and furthermore talks about E-Commerce Web Mining process. It likewise examines the different information wellsprings of Data Mining, Data Mining innovation and its applications in E-Commerce. It additionally calls attention to the extraordinary favourable circumstances of Web Mining innovation and it’s pretend in the use of E-Commerce.[5].Ren Jingbiao, Yin Shaohong, “Research and Improvement of Clustering Algorithm in Data Mining”, this paper examines the different ideas of Data Mining in detail. It additionally depicts the fundamental procedure of Data Mining and its different advances. The different ideas of Data Mining like the Clustering are likewise examined in this paper. It additionally

did the grouping examination calculation explore in light of the current Data Mining and concentrated on the current famous and ordinarily utilized clustering calculation i.e K-Means Algorithm.[10].Chintan Shah and Anjali Jivani, “Comparison of Data Mining Clustering Algorithms”, this paper demonstrates the investigation and Comparison between various Clustering Algorithms: K-means, Hierarchical, Density-Based, Grid-Based and so on. The different parameters taken for correlation are time and space many-sided quality and its different advantages and disadvantages and the product utilized for examination is WEKA (open source programming). From the examination, we presume that K-means is better calculation as looked at than others since it requires less investment than others.[12]. Yonghua Zhao and Hong Lin, “WEB data mining applications in e-commerce” the authors of this paper tries to combine the Web Data Mining & E-commerce which introduces the concept of web data mining in e-commerce and also describes the web data mining applications in e-commerce.[14].

6 Conclusion and Future Scope

From the above study, it can be concluded that in order to make the data structured it is important to apply Data-Mining on it. It can also be concluded that K-means Algorithm is better from all other clustering algorithms. In future; the comparison of the various Clustering Algorithms can be done on various other software’s like IBM SPSS Clementine, R, SPSS Statistics etc.Further, more review of the literature can be done for better understanding of Data Mining and its various applications.

7 References

- [1]. Guangcan yu, chuanlong xia, xingyue guo. “Research on Web Data Mining and its Application in Electronic-Commerce” IEEE 2009.
- [2]. Zhiwu Liu, Li Wang, “Study of Data Mining Technology Used for E-commerce” 2010 IEEE.
- [3]. S. P. Deshpande, V. M. Thakare. “Data mining system and applications: A review” IJDPS 2010.
- [4]. Limei Zhang, “Data mining application in customer relationship management” ICCASM 2010.
- [5]. Ren Jingbiao, Yin Shaohong, “Research and Improvement of Clustering Algorithm in Data Mining”, ICSPS 2010.
- [6]. “Business and data analytics: New innovations for the management of e-commerce” ELSEVIER 2012.
- [7]. Siti Khadijah Mohamad, Zaidatun Tasir, “Educational data mining: A review” ” ELSEVIER 2013.
- [8]. MD. Ezaz Ahmed, Preeti Bansal, “Clustering Technique on Search Engine Dataset using Data Mining Tool”, ACCT 2013.

- [9]. JOHN B. KILLORAN,” How to Use Search Engine Optimization Techniques to Increase Website Visibility” IEEE 2013.
- [10]. Chintan Shah and Anjali Jivani, “Comparison of Data Mining Clustering Algorithms”, NUICONE 2013.
- [11]. Baljeet Kaur, Sushila Madan, “A Fuzzy Expert System to Evaluate Customer’s Trusting B2C E-Commerce Websites” ICCSGD 2014.
- [12]. Yonghua Zhao, Hong Lin, “WEB data mining applications in e-commerce” ICCSE 2014.
- [13]. Rana Alaa El-Deen Ahmed,M.Elemam.Shehab, Shereen Morsy ,Nermeen Mekawie,“ Performance study of Classification Algorithms for consumer Online shopping attitudes and behaviour using Data Mining”, ICCNST 2015.
- [14]. Dr. Sanjay Kumar Dwivedi, Bhupesh Rawat, “A Review Paper on Data Pre-processing: A Critical Phase in Web Usage Mining Process”, IEEE 2015.
- [15]. Mohammad Salehan, DanJ.Kim, “Predicting the performance of online consumer reviews: A sentiment Mining approach to big data analytics” ELSEVIER 2015.
- [16]. Mohammad Noor Injadat, Fadi Salo, Ali Bou Nassif. “Data mining techniques in social media: A survey”, ELSEVIER 2016.
- [17]. Yan Li, Manoj A. Thomas, Kweku-Muata Osei-Bryson ,“A snail shell process model for knowledge discovery via data analytics”, ELSEVIER 2016.
- [18]. Hela Ltifi ,Emna Benmohamed ,Christophe Kolski ,Mounir Ben Ayed ,“Enhanced visual data mining process for dynamic decision-making” ,ELSEVIER 2016.