International Journal of Innovations & Research Analysis (IJIRA)
ISSN :2583-0295, Impact Factor: 5.449, Volume 02, No. 02, April- June, 2022,pp 31-35

USE OF DATA MINING IN E-COMMERCE PLATFORMS IN INDIA (A LITERATURE REVIEW)

Dr. Sarika Agarwal*
Dr. Mukesh Agarwal**

ABSTRACT

The development of the E-Commerce industry has been phenomenally high. However, its development is dependent on several components and the most significant of them is the use of data mining techniques and tools. In this research paper, we analysis of a literature review about it.

Keywords: *Data Mining, E-Commerce, Web Analysis and Reviews of Literature.*

Introduction

Data mining can help organizations identify the information that would be most suitable to put on the Web by analyzing competitors and potential customers, and determining the expertise of the organization. For any organization to get onto the World Wide Web, it needs to decide what information to put on its website. Data mining can also be used to assist organizations waning to conduct e-business with other organizations by providing information concerning selecting the best partners, identifying competitors, and determining the best pricing policies for its products or services. Summaries the E-Commerce and e-business applications of data mining as being the following:

- Analyzing customer profiles.
- Selecting a partner organization
- Assessing the similarity of user browsing patterns.
- Determining which products to market online.
- Categorization of web pages based on content.
- Assessment of the similarity of web page contents.
- Identifying web pages that are viewed together.

The website will assist online users in their decision-making by listing for instance other books that have been bought by people who also bought the particular book that the online user is currently looking at. One of the most useful applications of data mining in an E-Commerce environment can be found on the Amazon.com website. The website also makes use of data mining to bundle books that are often bought in pairs, together and to sell or market these at a reduced price if they are bought together. Amazon.com also uses data mining to profile its customers, and online users that return to the website after having bought a book, for instance, will be provided with lists of other book titles that they might be interested in based on the category of book that they just bought.

* Postdoctoral Fellow (Under PDF of ICSSR), Department of Management Studies, School of Studies in Management & Commerce, Guru Ghasidas University, (Central University of Chhattisgarh), Koni, Bilaspur, Chhattisgarh, India.
** Assistant Professor, Department of Commerce, School of Studies in Management & Commerce, Guru Ghasidas University, (Central University of Chhattisgarh), Koni, Bilaspur, Chhattisgarh, India.

Implementation Factors

The success of a DM exercise is driven to a very large extent by the following factors:

- **Reliability of the Data Available:** Although a given terabyte database may have hundreds of attributes per relation, the DM algorithms run on this dataset may be rendered defunct if the data itself was generated by manual and error-prone means and wrong default values were set. Also, the lesser the integration with legacy applications, the better the accuracy of the dataset.
- **Availability of Data with Rich Descriptions:** This means that unless the relations captured in the database are of a high degree, extracting hidden patterns and relationships among the various attributes will not make any practical sense. Availability of a large volume of data: This is mostly mandated for the statistical significance of the rules hold. The absence of say, at least a hundred thousand transactions will most likely reduce the usefulness of the rules generated from the transactional database.
- **Ease of Interfacing with Legacy Systems:** It is commonplace to find large organizations run on several legacy systems that generate huge volumes of data. A DM exercise which is usually preceded by other exercises like extract, transformation, and loading (ETL), data filtering, etc, should not add more overheads to system integration.
- **Ease of Quantification of the Return on Investment (ROI) in DM:** Although the earlier three factors may be favorable unless a strong business case can be easily made, investments in the next level of DM efforts may not be possible. In other words, the utility of the DM exercise needs to be quantified vis-a-vis the domain of application.

Reviews of Literature

Cao, Li & Yu (2011) narrates E-Commerce related data mining as a system to integrate statistical aspects, databases, as well as AI or artificial intelligence altogether for forming a novel concept or a holistic technique so that effective decision making can be possible. Data mining is a beneficial addition as far as online commerce promotion is concerned. At present, the implementation of data mining in the field of E-Commerce has turned into an attractive deal for businesses.

Liu and Wang, (2008) present data mining as a crucial tool to reposition online shopping organizations so that the enterprise can be supported through adequate data for commercial aspects. Currently, maximum organizations use E-Commerce tools with a huge amount of big data in the data storage. Only the data mining process can optimally use the huge scale of data for increasing the decision-making it enabling commercial intelligence.

Pan (2011) asserts that the major origin of data concerning online commercial organizations is the webpage. So, it's necessary to realize the functional process of Search engines so that the speed, the procedure, and the alterations of the aspects can be understood. For finding novel data Spiders, which are the software programs that are processed by the search engines. These are also known as bots or crawlers. Spiders or bots are specific software programs that search engine applies for requesting pages as well as downloading them. However, surprising for a few individuals, it's done by the search engine through a link that exists in a previous website for finding a newer one and requesting a copy of that specific page for downloading to the server. Search engines do this for running the ranking algorithms and this is demonstrated on the result page. So, the most important thing for a search engine is downloading an accurate website copy. Proper visual and readability aspects must be maintained for an online commercial website for the respective algorithms used for the search engine database. Techniques are required for removing the undesirable data in an automated way so that the data mining can transform the data for providing the dependable sensible or effective output through the data mining algorithm.

Kohavi et al (2001) postulate a design that collaborates and integrates a data mining system in an online commercial business. The structure is represented in the underlying and it comprises three major elements: Business Data Definition, Customer Interaction as well as Analysis, these have connections with each other employing data transferring bridges. As per the scholars, this integration properly resolves several issues in association with horizontal data mining techniques. It also includes the massive amount of endeavor needed for the data-pre-processing before using it in the mining tool. All these efforts make the mining outcomes functional. Due to the neat and close integration of three architectural elements with each other, the system permits the automated confirmation of a data warehouse/storing under the Analyzing part. The shared metadata throughout these three parts makes the structure simplified. Furthermore, the efficacy of the knowledge discovery procedure is increased with

the help of rich mining algorithmic equations along with the Analysis techniques. The closely integrated structure and shared metadata make things easier for deploying outcomes, as well as for closing the loop efficiently. Lastly, scholars demonstrated many issues and hardships which must be studied for enhancing this architectural design.

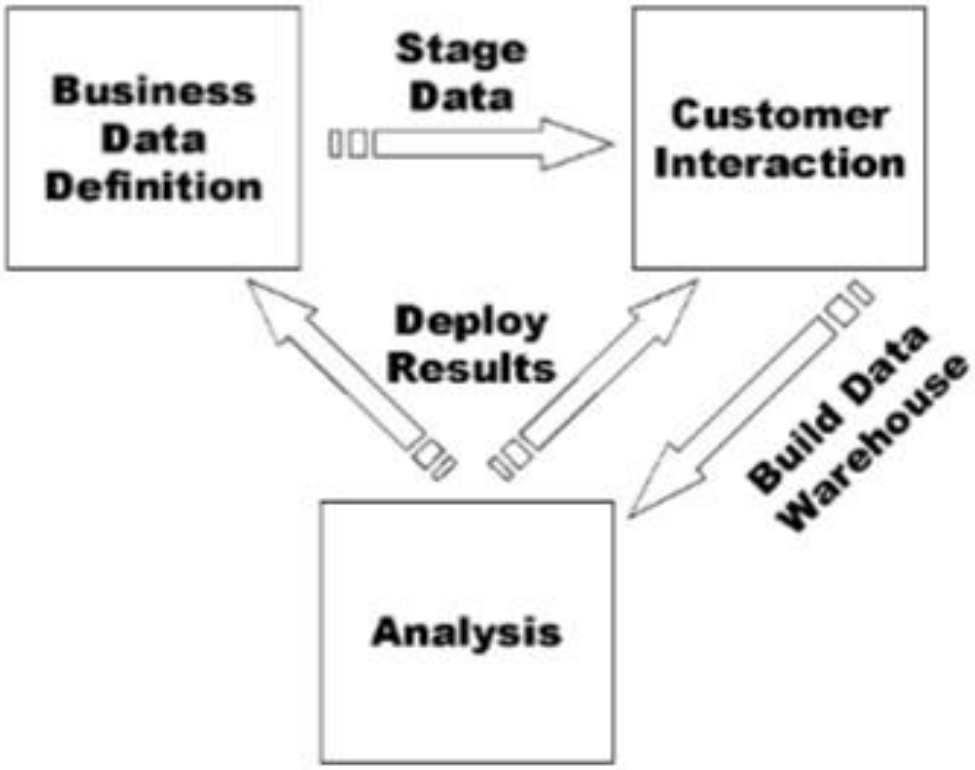

Figure 1: High-Level System Architecture

Frank et al (2003) devised a bare-bones part of an e-departmental store that integrates communicative mining functions. The online commerce system integrates the data mining algorithms into Microsoft SQL Server stored processes. The scholars have discovered the idea of utilizing the tools beyond data mining concerning particular supplier hunting, sales statistics, promotional aspects as well as email marketing. The close integration between the three elements of the structure permits the automated configuration of a data warehouse under the Analysis part. The disseminated information throughout the three elements, furthermore, simplifies this structure and is associated with the huge mining algorithmic equations and analysis tools further magnifying the efficacy of the knowledge revealing procedure. It initiates a better beginning for integrating and aligning data mining into online commerce systems. Many modular steps or tiers result from a multi-tier or multi-step application prototype. Every single modular tier can be situated on a separate computer machine. The data tier contains all the information required to do the application. Usually, these data are saved in a database. The middle tier in this multi-tier setup works like a mediator that connects the information compiled in a data tier with the application users. Each user asks for information, (for viewing the catalog of items) and passes through the middle tier before obtaining the database. Similarly, feedback to requests for data travels utilizing the middle tier before going to the user. The middle tier applies the commercial logic as well as presentation logic for controlling communications between information and users. Business or trade logic uses business law (for example, a promotion scheme) and is utilized for ensuring effective data before updating the database or retrieving it for the user. The third tier conforms to the client tier using the user interface for the application. Users, as well as system administrators, communicate directly with the client tier with the help of the user interface. When it comes to the shopping system, the client becomes an HTML and WML-enabled web browser. The user interface is also used by the user or administrator when it comes to the client tier.

Hong Yu et al (2004) recommend a general suggestion to collect as well as mine data from an online commercial system. This recommended structure has 4 major modules, namely, Data Collection, Data Pre-processing, Pattern Discovery, as well as Knowledge Analysis. The integrated structure can fulfill the requirements of knowledge administration as far as E-Commerce is concerned.

Conclusion

After reviewed of various studies, we found that the challenges, which had to be faced to the interface, data collection & preprocessing, pattern discoveries, data mining techniques, and risk analysis in the E-Commerce industry have been studied mainly in the various research work done by various scholars in the past. Or, they have only reviewed the facts that helped the E-Commerce industries to meet the challenges posed by consumer trust.

At present, the implementation of data mining in the field of E-Commerce has turned into an attractive deal for businesses. So, in our study, we have tried to test, consumer trust in an E-Commerce business related to perceived security, reputation & store network privacy, the influence of consumer trust on perceived risk, and the relationship between the consumer purchase intention and consumer perceived risk in regards with an E-Commerce business.

References

1. Agrawal and Agrawal (2014). Review on Data Mining Tools, IJISET - International Journal of Innovative Science, Engineering & Technology. 2.
2. Awais, M. and Samin, T. (2012), “Advanced SWOT Analysis of E-Commerce”, IJCSI International Journal of Computer science Issues, Vol 9, Issue 2, No 2, pp. 569-574
3. Bhat, D. S. Kansana, K. & Majid, J. (2016). A Review Paper on E-Commerce.
4. Bhatnagar, Jadye, and Nagar (2012). Data Mining Techniques & Distinct Applications: A Literature Review, International Journal of Engineering Research & Technology. 2.
5. Bo, P. Lillian, L. and Shivakumar, V. (2002). “Thumbs up? Sentiment classification using machine learning techniques”, In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 79– 86.
6. Brown, (2015). “What sells most in the Indian E-Commerce market”. Retrieved from http://indianonlineseller.com/2017/02/what-sells-most-in-the-indianE-Commerce-market/ accessed on 26th July 2017 at 1:15 am
7. Chien-Liang, L. Wen-Hoar, H. Chia-Hoang, L. Gen-Chi, L. And Emery, J. (2012). Movie Rating And Review Summarizationin Mobile Environment”, IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews. 42.
8. Chou, D. C. and Chou, A. Y. (2000). “The E-Commerce Revolution, A Guide to the Internet Revolution in Banking” information systems management. 51-57
9. D’silva, B. D’Silva, S. and Bhuptani, R. S. K. (2010), “Behavioral Aspect of Teenagers Towards Internet Banking: An empirical study”, Indian journal of marketing, 40.10, pp.44-53
10. Das, D. K. & Ara, A. (2015, July). Growth of E-Commerce in India. International Journal Of Core Engineering & Management. 2(4), 25-33.
11. Dasgupta, P. & Sengupta, K. (2002). E-Commerce in the Indian Insurance Industry. Electronic Commerce Research. 2, 43-60.
12. De Projeto (2014). The Impact of Channel Conflict on Channel Efficiency.
13. Dnyanesh G. R. And Singh, S. (2014). ” An Ontology-Based Text Mining Method To Develop D-Matrix From Unstructured Text”, IEEE Transactions On Systems, Man, And Cybernetics: Systems. 44.
14. Dogra and Wala (2015). A Review Paper on Data Mining Techniques and Algorithms, International Journal of Advanced Research in Computer Engineering & Technology. 4(5).
15. Dokania and Kaur (2018). International journal of engineering sciences & research technology, International Journal of Data Mining Techniques and Applications. 2
16. Donald E. Brown, (2016). ” Text Mining The Contributors To Rail Accidents”, IEEE Transactions On Intelligent Transportation Systems. 17.
17. Hilage and Kulkarni (2012). Review of literature on data mining, International Conference on Issues and Challenges in Intelligent Computing Techniques. 8.

18. Hsieh, C. (2005). “Implementing Self-Service Technology To Gain Competitive Advantages” Communications of the IIMA. 1, 77-83.

19. Jackie Gilbert Bette, A. S. (2001). Ethical Issues in electronic Commerce: Journal of Business Ethics. 34, 75-85.

20. Paris, D. Bahari, M. Iahad, N. & Ismail, W. (2016). Systematic literature review of E-Commerce implementation studies. 89, 422-438.

21. Rina, D. (2016). Challenges and Future Scope of E-Commerce in India. International Journal of Emerging Trends & Technology in Computer Scienc. 5(2), 232-235.

22. Rust, R. T. and Chung, T. S. (2006). “Marketing Models of Service and Relationships” Marketing science. 25, 560-580.

23. Singh, O. P. and Lalwani, R. (2007). “Internet Banking-A Comparative Study on Public and Private Sector Banks in India”, Indian journal of public enterprise. 42, 51-60.

24. Tsay and Agrawal (2004). Modeling Conflict and Coordination in Multi-Channel Distribution Systems: A Review, Handbook of Quantitative Supply Chain Analysis. 557-606.

25. Vaithianathan, S. (2010). A review of E-Commerce literature on India and research agenda for the future. Electron Commer Res 10, 83–97.

26. Xiuzhen, Z. Lishan, C. And Wang, Y. (2014). “Commtrust: Computing Multi-Dimensional Trust By Mining E-Commerce Feedback Comments”, IEEE Transactions On Knowledge And Data Engineering. 26.

27. Yue, H. and Xiaojun, W. (2014). ” Ppsgen: Learning-Based Presentation Slides Generation For Academic Papers”, IEEE Transactions On Knowledge And Data Engineering. 27.

28. Zabihollah Rezaee Kenneth R. Lambert and W. Ken Harmon (2006). electronic commerce Education: Analysis of existing courses; Accounting Education: An International Journal. 73-88.