

Study of different Data Mining system & Platform

Neha Rathee¹, Sarika Choudhary²

¹ Student, CBS Group of Institutions, Haryana, India

² Assistant. Professor, CBS Group of Institutions, Haryana, India

Abstract- This Paper gives an introduction of Data Mining System and development of different data mining systems. How many data mining tools or systems really exist? What are different platform under which these tools or systems has been designed? What are different areas for which these data mining system has been devised or developed. After doing study and research on data mining system. we have to decide which platform is best for the development of Data Mining System and to understand the importance of data mining system and platform on which it has been developed.

As we know what is importance of data mining system for big and large organization, when the company is working in different parts of country and producing large amount of data then it very important to analyze that data which account for productivity and cost benefits from that data. Now it is important to understand that which part of the organization is giving benefit and where the organization is losing. So to take any decision, by the help of Data Mining tools the overall data is combined and given as input to produce as output in the form of high level managerial decision so that decision can increase the profit of organization. Profit of the organization can be seen in terms of working efficiency, productivity etc.

Keywords: Data Mining, Data Mining System , Platform

INTRODUCTION

Data Mining is the process or it's a way to of extract information from large data sets or group of data through the use of algorithms and techniques.

Data mining is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules.

Definition

Data mining is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules [1].

Data mining is an interdisciplinary subfield of computer science which involves computational process of large data sets' patterns discovery. The goal of this advanced analysis process is to extract information from a data set and transform it into an understandable structure for further use. The methods used are at the juncture of artificial intelligence, machine learning, statistics, database systems and business intelligence. Data Mining is about solving problems by analyzing data already present in databases [2].

Data mining consists of five major elements:

1. Extract, transform, and load transaction data onto the data warehouse system.
2. Store and manage the data in a multidimensional database system.
3. Provide data access to business analysts and information technology professionals.
4. Analyze the data by application software.
5. Present the data in a useful format, such as a graph or table.

Data Mining System

Data mining refers to the process of extracting new and useful knowledge from large amounts of data [4]. Data

mining is widely used to solve many business problems, such as customer profiling [5], customer behavior modeling [6], credit scoring, product recommendation, direct marketing [7], cross selling, fraud detection [8, 9]. Data mining is adopted in many industries, e.g., retail [10], bank, finance [11], and so on [12, 13].

The solution to a data mining problem is carried out in the data mining process, which varies depending on the application domain. In general, a data mining process consists of the following seven steps [14]:

1. Identify the business problems.
2. Identify and study data sources, and select data.
3. Extract and preprocess data.
4. Mine the data, e.g., discover association rules or build predictive models.
5. Verify the mining results.
6. Deploy models in the business process.
7. Measure the return on investment (ROI).

Different Types

KNIME

KNIME is the leading open platform for data-driven innovation helping organizations to stay ahead of change. Use our open-source, enterprise-grade analytics platform to discover the potential hidden in your data, mine for fresh insights or predict new futures [15].

ADaM

The Algorithm Development and Mining System (ADaM) developed by the Information Technology and Systems Center at the University of Alabama in Huntsville is used to apply data mining technologies to remotely-sensed and other scientific data. The mining and image processing toolkits consist of interoperable components that can be linked together in a variety of ways for application to diverse problem domains. ADaM has over 100 components that can be configured to create customized mining processes [16].

AlphaMiner

AlphaMiner is developed by the E-Business Technology Institute (ETI) of the University of Hong Kong under the support from the Innovation and Technology Fund (ITF) of the Government of the Hong Kong Special Administrative Region (HKSAR). It is an open source data mining platform that provides the best cost-and-performance ratio for data mining applications [17].

FUTURE SCOPE

By Study we come to know that only little system has been designed in the area of Data mining using some open source technology. So we can focus on this area because of its many more benefits of using open source technology.

REFERENCES

1. Xingquan Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", ISBN 978-1-59904-252, Hershey, New York, 2007.
2. Joseph, Zernik, "Data Mining as a Civic Duty – Online Public Prisoners Registration Systems", International Journal on Social Media: Monitoring, Measurement, Mining, vol. - 1, no.-1, pp. 84-96, September 2010.
3. Witten IH, Frank E. Data mining: practical machine learning tools and techniques with Java implementations. 2nd edition. San Francisco (CA): Morgan Kaufmann; 2005.
4. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann(2000).
5. Adomavicius, G., Tuzhilin, A.: Using data mining methods to build customer profiles. Computer (2001).

6. Bounsaythip, C., Rinta, E.: Overview of data mining for customer behavior modeling. Technical report, VTT Information Technology (2001).
7. Ling, C.X., Li, C.: Data mining for direct marketing: Problems and solutions. American Association for Artificial Intelligence (1998).
8. Rygielski, C., Wang, J.C., C, D.: Data mining techniques for9. Apte, C., Liu, B., Pednault, E.P.D., Smyth, P.: Business applications of data mining. Communications of the ACM 45 (2002) 483-502.
9. Ahmed, S.R.: Applications of data mining in retail business. In: Proceedings of the International Conference on Information Technology: Coding and Computing.(2004).
10. Kovalerchuk, B., Vityaev, E.: Data Mining in finance: Advances in Relational and Hybrid Methods. Kluwer Academic Publishers (2000).
11. Han, J., Altman, R.B., Kumar, V., Mannila, H., Pregibon, D.: Emerging scientific applications in data mining. Communications of the ACM 45 (2002) 54-58.
12. Grossman, R., Kamath, C., Kegelmeyer, P., Kumar, V., Namburu, R.: Data Mining for Scientific and Engineering Applications. Kluwer Academic Publishers (2001).
13. Huang, J.: Data mining overview. Technical report, E-Business Technology Institute (2006)
14. Goebel, M., Gruenwald, L.: A survey of data mining and knowledge discovery software tools. In: SIGKDD Explorations. Volume 1., ACM SIGKDD (1999) 20-33.

15. <https://www.knime.org/>

16. <http://projects.itsc.uah.edu/datamining/adam/>

17. <http://www.eti.hku.hk/alphaminer/>

BIOGRAPHIES



Neha Rathee, Student of Mtech in Computer Science Engineering from CBS Group of Institutions. Started research work by using Data Mining Platform.



Sarika Choudhary, Assistant Professor in Computer Science Engineering Department of CBS Group of Institutions.