

# Study and Result of Data Mining system using Open Source

<sup>1</sup>Neha Rathee, Sarika Choudhary<sup>2</sup>

<sup>1</sup>M.Tech Student(CSE) CBS Group of Institutions, Fathepuri Jhajjar

<sup>2</sup>Asst. Prof. in CSE Dept. , CBS Group of Institutions, Fathepuri Jhajjar

## Abstract

This Paper gives result of the developed Data Mining System using Open Source. We have developed the data mining system under two areas i.e "Product" & "Share". For the category "Product", we are analyzing Product under three different sub-category i.e, 'Product Purchased', 'Customer Points', 'Customer Bills' & for the category "Share", we are analyzing Share under three different sub-category i.e, 'Share in Demand', 'Share Price'.

**Keywords:** Data Mining, Data Mining System

## 1.Introduction

Data mining refers to extracting or 'mining' interesting knowledge from large amounts of data [1]. It provides a means of extracting previously unknown, predictive information from the base of accessible data in data warehouses. Data mining tools use sophisticated, automated algorithms to discover hidden patterns, correlations, and relationships among organizational data. These tools are used to predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions [2].

Data Mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

### Process

The process of data mining consists of three stages:

1. The initial exploration.
2. Model building or pattern identification with validation/verification.
3. Deployment (i.e., the application of the model to new data in order to generate predictions).

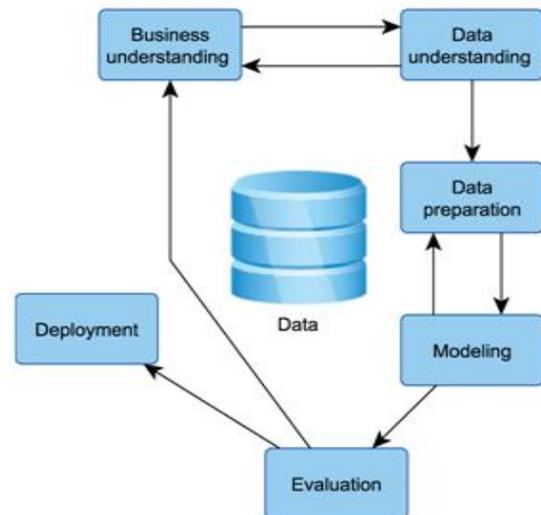


Figure : 1, Diagram of Data Mining Process

### Stage 1: Exploration.

This stage usually starts with data preparation which may involve cleaning data, data transformations, selecting subsets of records and - in case of data sets with large numbers of variables ("fields") - performing some preliminary feature selection operations to bring the number of variables to a manageable range (depending on the statistical methods which are being considered). Then, depending on the nature of the analytic problem, this first stage of the process of data mining may involve anywhere between a simple choice of straightforward predictors for a regression model, to elaborate exploratory analyses using a wide variety of graphical and statistical methods (see Exploratory Data Analysis (EDA)) in order to identify the most relevant variables and determine the complexity and/or the general nature of models that can be taken into account in the next stage.

### Stage 2: Model building and validation.

This stage involves considering various models and choosing the best one based on their predictive performance (i.e., explaining the variability in question and producing stable results across samples). This may sound like a simple operation, but in fact, it sometimes involves a very elaborate process. There are a variety of techniques developed to achieve that goal - many of which are based on so-called "competitive evaluation of models," that is, applying different models to the same data set and then comparing their performance to choose

the best. These techniques - which are often considered the core of predictive data mining - include: Bagging (Voting, Averaging), Boosting, Stacking (Stacked Generalizations), and Meta-Learning.

### Stage 3: Deployment.

That final stage involves using the model selected as best in the previous stage and applying it to new data in order to generate predictions or estimates of the expected outcome.

The concept of Data Mining is becoming increasingly popular as a business information management tool where it is expected to reveal knowledge structures that can guide decisions in conditions of limited certainty. Recently, there has been increased interest in developing new analytic techniques specifically designed to address the issues relevant to business Data Mining (e.g., Classification Trees), but Data Mining is still based on the conceptual principles of statistics including the traditional Exploratory Data Analysis (EDA) and modeling and it shares with them both some components of its general approaches and specific techniques.

### Problem Statement

In this research work, we aimed to develop a Data Mining application for two areas i.e for Product and Share. It was aimed to develop an Open Source based Data Mining Application .The Developed Application will work on any of the browser and the user of the application can predict which product is under requirement and which is not under requirement, which product is in most demanded by the consumer who are coming for purchase, which customer is having the maximum points and who is having the minimum one, which share is in demand and which share is of high coast .

### Result

The developed Data Mining system using Open Source, which can be shown in the below figure. By the use of developed system the user can take managerial decision , which is ultimate goal of any Developed Data Mining System.



Figure : 2, Gateway of Data Mining System

## 2.Conclusion

In this research work, we explored the problem of Data Mining that how we can store the data, how to retrieve record on front so that it can be proven beneficial to take any managerial decision.

The Developed data mining system will decide which product is in demand & according to which the data is mined from the stored data for that product, what is the actual profit & loss, who is the competitor of this product, what wrong , where the company or industry is lacking, what to do to increase the consumption, how to launch any product. It will be very helpful to industry who is interested in doing market analysis in the particular field. It is very fast in processing and user using that developed system can get the result quickly. The developed system can be used for online information extraction by implementing the developed data mining system online.

## References

- [1]. The Data Mine: [www.the-data-mine.com](http://www.the-data-mine.com)
- [2]. KDnuggets - Data Mining, Web Mining, and Knowledge Discovery Guide: [www.kdnuggets.com](http://www.kdnuggets.com)
- [3]. Xingquan Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", ISBN 978-1-59904-252, Hershey, New York, 2007.
- [4]. 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.
- [5]. Witten IH, Frank E. Data mining: practical machine learning tools and techniques with Java implementations. 2nd edition. San Francisco (CA): Morgan Kaufmann; 2005.
- [6]. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann(2000).
- [7]. Adomavicius, G., Tuzhilin, A.: Using data mining methods to build customer profiles. Computer (2001).
- [8]. Bounsaythip, C., Rinta, E.: Overview of data mining for customer behavior modeling. Technical report, VTT Information Technology (2001).
- [9]. Ling, C.X., Li, C.: Data mining for direct marketing: Problems and solutions. American Association for Artificial Intelligence (1998).
- [10]. 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.
- [11]. Ahmed, S.R.: Applications of data mining in retail business. In: Proceedings of the International Conference on Information Technology: Coding and Computing.(2004).
- [12]. Kovalerchuk, B., Vityaev, E.: Data Mining in finance: Advances in Relational and Hybrid Methods. Kluwer Academic Publishers (2000).

- [13]. 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.
- [14]. Grossman, R., Kamath, C., Kegelmeyer, P., Kumar, V., Namburu, R.: *Data Mining for Scientific and Engineering Applications*. Kluwer Academic Publishers (2001).
- [15]. Huang, J.: *Data mining overview*. Technical report, E-Business Technology Institute (2006)
- [16]. Goebel, M., Gruenwald, L.: A survey of data mining and knowledge discovery software tools. In: *SIGKDD Explorations*. Volume 1., ACM SIGKDD (1999) 20-33.
- [17]. <https://www.knime.org/>
- [18]. <http://projects.itsc.uah.edu/datamining/adam/>
- [19]. <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.