

A Novel Approach to the Overview on the Data Mining Techniques

Amos R¹, Vinay M², Rudresh Y R³

¹Assistant Professor, Department of MCA, MIT Mysore.

² Assistant Professor, Department of Computer Science, Sesadripuram Degree College, Mysore .

³ Assistant Professor, Department of Computer Science, SBRR Mahajana First Grade College, Mysore

Abstract:-

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data mining depends on effective data collection, warehousing, and computer processing. Data mining processes are used to build machine learning models that power applications including search engine technology and website recommendation programs. Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data mining depends on effective data collection, warehousing, and computer processing. Data mining processes are used to build machine learning models that power applications including search engine technology and website recommendation programs. Data mining involves exploring and analyzing large blocks of information to glean meaningful patterns and trends. It can be used in a variety of ways, such as database marketing, credit risk management, fraud detection, spam Email filtering, or even to discern the sentiment or opinion of users. The data mining process breaks down into five steps. First, organizations collect data and load it into their data warehouses. Next, they store and

manage the data, either on in-house servers or the cloud. Business analysts, management teams and information technology professionals access the data and determine how they want to organize it. Then, application software sorts the data based on the user's results, and finally, the end-user presents the data in an easy-to-share format, such as a graph or table.

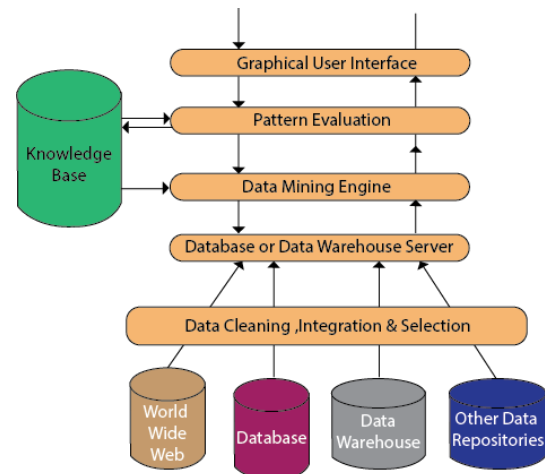
INTRODUCTION

Along with the above example, the era of Big Data has arrived. Every day, 2.5 quintillion bytes of data are created and 90 percent of the data in the world today were produced within the past two years. Our capability for data generation has never been so powerful and enormous ever since the invention of the information technology in the early 19th century. As another example, on 4 October 2012, the first presidential debate between President Barack Obama and Governor Mitt Romney triggered more than 10 million tweets within 2 hours. Among all these tweets, the specific moments that generated the most discussions actually revealed the public interests, such as the discussions about medicare and vouchers. Such online discussions provide a new means to sense the public interests and generate feedback in realtime, and are mostly appealing compared to generic media, such as radio or TV broadcasting. Another example is Flickr, a public picture sharing site, which received 1.8 million photos per day, on average, from February to March 2012. Assuming the size of

each photo is 2 megabytes (MB), this requires 3.6 terabytes (TB) storage every single day. Indeed, as an old saying states: “a picture is worth a thousand words,” the billions of pictures on Flickr are a treasure tank for us to explore the human society, social events, public affairs, disasters, and so on, only if we have the power to harness the enormous amount of data. The above examples demonstrate the rise of Big Data applications where data collection has grown tremendously and is beyond the ability of commonly used software tools to capture, manage, and process within a “tolerable elapsed time.” The most fundamental challenge for Big Data applications is to explore the large volumes of data and extract useful information or knowledge for future actions. In many situations, the knowledge extraction process has to be very efficient and close to real time because storing all observed data is nearly infeasible. For example, the square kilometer array (SKA) in radio astronomy consists of 1,000 to 1,500 15-meter dishes in a central 5-km area. It provides 100 times more sensitive vision than any existing radio telescopes, answering fundamental questions about the Universe. However, with a 40 gigabytes (GB)/second data volume, the data generated from the SKA are exceptionally large. Although researchers have confirmed that interesting patterns, such as transient radio anomalies can be discovered from the SKA data, existing methods can only work in an offline fashion and are incapable of handling this Big Data scenario in real time. As a result, the unprecedented data volumes require an effective data analysis and prediction platform to achieve fast response and real-time classification for such Big Data.

Data Mining Architecture

The significant components of data mining systems are a data source, data mining engine, data warehouse server, the pattern evaluation module, graphical user interface, and knowledge base.



Data Mining Engine:

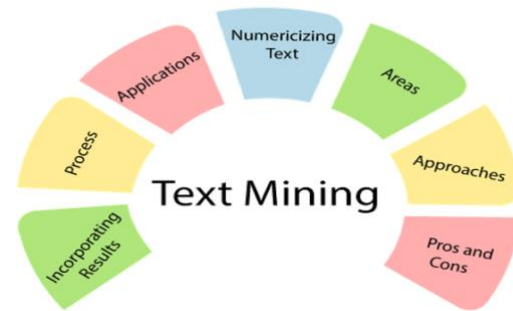
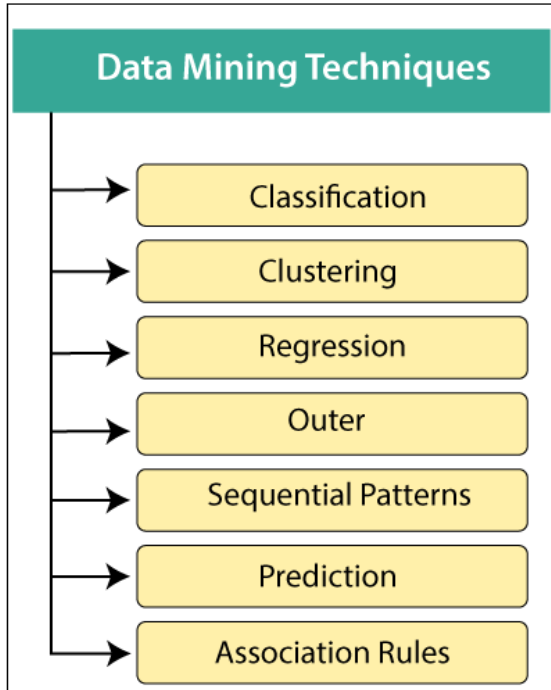
The data mining engine is a major component of any data mining system. It contains several modules for operating data mining tasks, including association, characterization, classification, clustering, prediction, time-series analysis, etc. In other words, we can say data mining is the root of our data mining architecture. It comprises instruments and software used to obtain insights and knowledge from data collected from various data sources and stored within the data warehouse.

Data Mining Techniques

Data mining includes the utilization of refined data analysis tools to find previously unknown, valid patterns and relationships in huge data sets. These tools can incorporate statistical models, machine learning techniques, and mathematical algorithms, such as neural networks or decision trees. Thus, data mining incorporates analysis and prediction. Depending on various methods and technologies from the intersection of machine learning, database management, and statistics, professionals in data mining have devoted their careers to better understanding how to process and make conclusions from the huge amount of data, but what are the methods they use to make it happen?

In recent data mining projects, various major data mining techniques have been developed and used, including association, classification,

clustering, prediction, sequential patterns, and regression.



The text mining market has experienced exponential growth and adoption over the last few years and also expected to gain significant growth and adoption in the coming future. One of the primary reasons behind the adoption of text mining is higher competition in the business market, many organizations seeking value-added solutions to compete with other organizations. With increasing completion in business and changing customer perspectives, organizations are making huge investments to find a solution that is capable of analyzing customer and competitor data to improve competitiveness.

Facebook Data Mining

In this digital era, the social platform has become inevitable. Whether we like this platform or not, there is no escape. Facebook allows us to interact with friends and family or to stay up to date about the latest stuff happening around the world. Facebook has made the world seems much smaller. Facebook is one of the most important sources of online business communication. The business holders make the most out of this platform. The most important reason for which this platform is most accessed is because of its characteristic of being the oldest video and photo sharing social media tool. A Facebook page helps the people to get aware of the brand through the media content shared. The platform supports the businesses to reach out to their audience and then establish their business belonging to Facebook usage itself.

Data Mining in Healthcare

Data Mining Steps

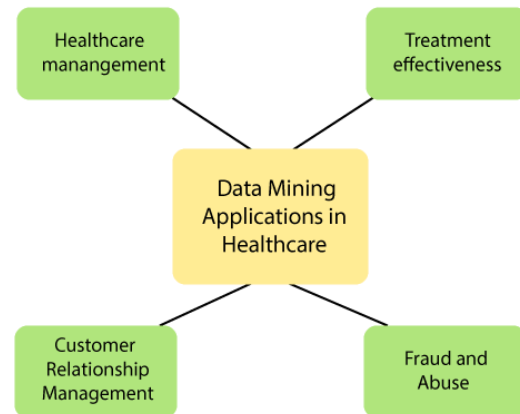
- Understand Business
- Understand the Data
- Prepare the Data
- Model the Data
- Evaluate the Data
- Deploy the Solution

Types of Data Mining:-

Text Data Mining

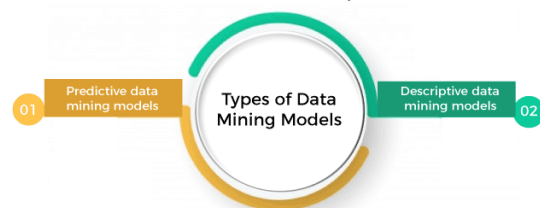
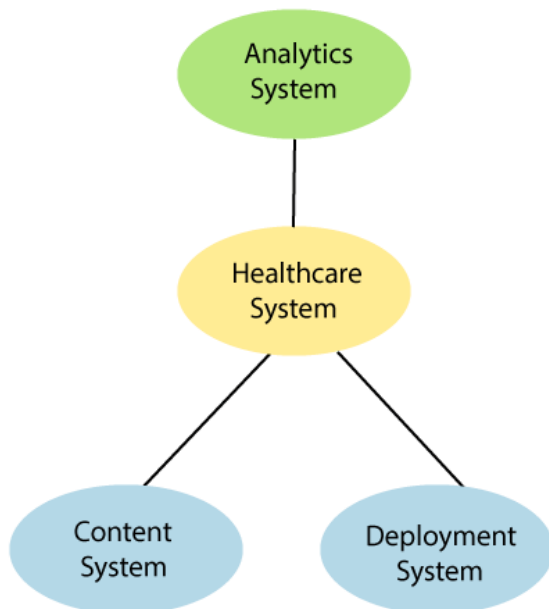
Text data mining can be described as the process of extracting essential data from standard language text. All the data that we generate via text messages, documents, emails, files are written in common language text. Text mining is primarily used to draw useful insights or patterns from such data.

Various sectors effectively use data mining. It enables the retail sectors to display customer response and helps the banking sector to predict customer profitability. It serves many similar sectors such as manufacturing, telecom, healthcare, automotive industry, education, and many more. Data mining holds incredible potential for healthcare services due to the exponential growth in the number of electronic health records. Previously Doctors and physicians hold patient information in the paper where the data was quite difficult to hold. Digitalization and innovation of new techniques reduce human efforts and make data easily assessable. For example, the computer keeps a massive amount of patient data with accuracy, and it improves the quality of the whole data management system. Still, the major challenge is what should healthcare services providers do to filter all the data efficiently? This is the place where data mining has proven to be extremely useful.



Data Mining Models

Data mining uses raw data to extract information and present it uniquely. The data mining process is usually found in the most diverse range of applications, including business intelligence studies, political model forecasting, web ranking forecasting, weather pattern model forecasting, etc. In business operation intelligence studies, business experts mine huge data sets related to a business operation or a market and try to discover previously unrecognized trends and relationships. Data mining is also used in organizations that utilize big data as a raw data source to extract the required data.

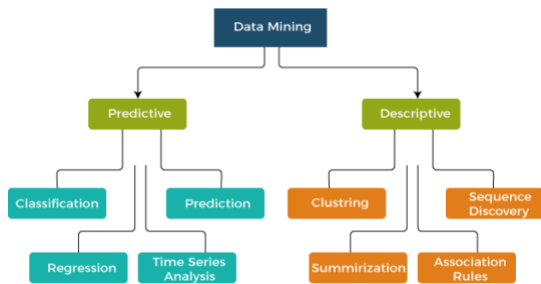


1. Predictive data mining models
2. Descriptive data mining models

Predictive data mining model:

A predictive data mining model predicts the values of data using known results gathered from the different data sets. Predictive modeling can not be classified as a separate discipline; it occurs in all organizations or industries across all

disciplines. The main objective of predictive data mining models is to predict the future based on the past data, generally but not always on the statistical modeling.



Descriptive data mining model

A descriptive model differentiates the patterns and relationships in data. A descriptive model does not attempt to generalize to a statistical population or random process. A predictive model attempts to generalize to a population or random process. Predictive models should give prediction intervals and must be cross-validated; that is, they must prove that they can be used to make predictions with data that was not used in constructing the model.

Descriptive analytics focuses on the summarization and conversion of the data into useful information for reporting and monitoring.

Advantages of Data Mining

Since we live and work in a data-centric world, it's essential to get as many advantages as possible. Data mining provides us with the means of resolving problems and issues in this challenging information age. Data mining benefits include

- It helps companies gather reliable information
- It's an efficient, cost-effective solution compared to other data applications

- It helps businesses make profitable production and operational adjustments
- Data mining uses both new and legacy systems
- It helps businesses make informed decisions
- It helps detect credit risks and fraud
- It helps data scientists easily analyze enormous amounts of data quickly
- Data scientists can use the information to detect fraud, build risk models, and improve product safety
- It helps data scientists quickly initiate automated predictions of behaviors and trends and discover hidden pattern

REFERENCES

- [1] R. Ahmed and G. Karypis, "Algorithms for Mining the Evolution of Conserved Relational States in Dynamic Networks," *Knowledge and Information Systems*, vol. 33, no. 3, pp. 603-630, Dec. 2012.
- [2] M.H. Alam, J.W. Ha, and S.K. Lee, "Novel Approaches to Crawling Important Pages Early," *Knowledge and Information Systems*, vol. 33, no. 3, pp 707-734, Dec. 2012.
- [3] S. Aral and D. Walker, "Identifying Influential and Susceptible Members of Social Networks," *Science*, vol. 337, pp. 337-341, 2012.
- [4] A. Machanavajjhala and J.P. Reiter, "Big Privacy: Protecting Confidentiality in Big Data," *ACM Crossroads*, vol. 19, no. 1, pp. 20-23, 2012.
- [5] S. Banerjee and N. Agarwal, "Analyzing Collective Behavior from Blogs Using Swarm Intelligence," *Knowledge and Information Systems*, vol. 33, no. 3, pp. 523-547, Dec. 2012.
- [6] E. Birney, "The Making of ENCODE: Lessons for Big-Data

- Projects," *Nature*, vol. 489, pp. 49-51, 2012.
- [7] J. Bollen, H. Mao, and X. Zeng, "Twitter Mood Predicts the Stock Market," *J. Computational Science*, vol. 2, no. 1, pp. 1-8, 2011.
- [8] S. Borgatti, A. Mehra, D. Brass, and G. Labianca, "Network Analysis in the Social Sciences," *Science*, vol. 323, pp. 892-895, 2009.
- [9] J. Bughin, M. Chui, and J. Manyika, *Clouds, Big Data, and Smart Assets: Ten Tech-Enabled Business Trends to Watch*. McKinsey Quarterly, 2010.
- [10] D. Centola, "The Spread of Behavior in an Online Social Network Experiment," *Science*, vol. 329, pp. 1194-1197, 2010.
- [11] E.Y. Chang, H. Bai, and K. Zhu, "Parallel Algorithms for Mining Large-Scale Rich-Media Data," *Proc. 17th ACM Int'l Conf. Multimedia, (MM '09)*, pp. 917-918, 2009.