Process of Data Mining and Its Benefits

Dr.R.Praba[1], S.Guruvigneswari[2], V.Karthika,K.Swathi[3]

[1]*Assistant Professor, Department of Information Technology, Dr.N.G.P Arts and Science College, Coimbatore, India*

[2,3]*Student, Department of Information Technology, Dr.N.G.P Arts and Science College, Coimbatore, India*

Abstract: Data mining is a field of intersection of computer science and statistics used to discover patterns in the information bank. The main aim of the data mining process is to extract the useful information from the dossier of data and mold it into an understandable structure for future use.

Data mining may be regarded as the process of discovering insightful and predictive models from massive data. It is the art of extracting useful information from large amounts of data. It combines traditional data analysis with sophisticated algorithms for processing large amount of data.

Keywords: Data mining, Uses, Benefits, Process, Techniques, Applications

DATA MINING

Data mining is the process of extracting patterns and other useful information from large data sets. It's sometimes known as knowledge discovery in data or KDD. Thanks to the rise of big data and advancements in data warehousing technologies, the use of data mining techniques has grown in recent decades, turning raw data into valuable knowledge that companies can use.

Data mining has improved corporate decision-making through clever data analytics. Data mining techniques can be broadly classified into two categories:

- Defining the target dataset
- Forecasting outcomes using machine learning methods

USES OF DATA MINING

The following are some of the applications of data mining:

- To achieve a corporate goal
- To answer business or research questions
- To contribute to problem-solving
- To aid in the accurate prediction of outcomes
- To analyse and predict trends and anomalies
- To inform forecasts
- To identify gaps and mistakes in processes, such as supply chain bottlenecks or incorrect data entry

BENEFITS OF DATA MINING

The following are some of the benefits of data mining:

- It assists businesses in gathering reliable data
- It assists organisations in making well-informed decisions
- It is a time- and cost-effective solution when compared to other data applications
- It enables organisations to make cost-effective production and operational changes
- It aids in the detection of credit issues and fraud
- It enables data scientists to quickly evaluate massive amounts of data. Data scientists can then use the data to spot fraud, create risk models, and improve product safety

It enables data scientists to create behaviour and trend forecasts and uncover hidden patterns.

DATA MINING PROCESS

Step 1 : Understanding the buisness

Before any data is touched, extracted, cleaned, or analyzed, it is important to understand the underlying entity and the project at hand. What are the goals the company is trying to achieve by mining data? What is their current business situation? What are the findings of a SWOT analysis? Before looking at any data, the mining process starts by understanding what will define success at the end of the process.

Step 2 : Understand the data

Once the business problem has been clearly defined, it's time to start thinking about data. This includes what sources are available, how they will be secured and stored, how the information will be gathered, and what the final outcome or analysis may look like. This step also includes determining the limits of the data, storage, security, and collection and assesses how these constraints will affect the data mining process.

Step 3 : Prepare the data

Data is gathered, uploaded, extracted, or calculated. It is then cleaned, standardized, scrubbed for outliers, assessed for mistakes, and checked for reasonableness. During this stage of data mining, the data may also be checked for size as an oversized collection of information may unnecessarily slow computations and analysis.

Step 4 : Build the model

With our clean data set in hand, it's time to crunch the numbers. Data scientists use the types of data mining above to search for relationships, trends, associations, or sequential patterns. The data may also be fed into predictive models to assess how previous bits of information may translate into future outcomes.

Step 5 : Evaluate the results:

The data-centered aspect of data mining concludes by assessing the findings of the data model or models. The outcomes from the analysis may be aggregated, interpreted, and presented to decision-makers that have largely been excluded from the data mining process to this point. In this step, organizations can choose to make decisions based on the findings.

Step 6 : Implement change and monitor:

The data mining process concludes with management taking steps in response to the findings of the analysis. The company may decide the information was not strong enough or the findings were not relevant, or the company may strategically pivot based on findings. In either case, management reviews the ultimate impacts of the business and recreates future data mining loops by identifying new business problems or opportunities.

DATA MINING TECHNIQUES

Tracking patterns:

Learning to discover patterns in your data sets is one of the most basic data mining techniques. This is frequently an identification of a periodic anomaly in the data or the ebb and flow of a particular variable through time.

For example, you may find that sales of a particular product increase immediately before the holidays, or that warmer weather sends more visitors to your website.

Forecasting:

Prediction is one of the most important data mining methods, as it's used to forecast the types of data you'll see in the future. In many circumstances, simply noticing and understanding previous patterns is sufficient to provide a reasonable prediction of what will occur in the future. For example, you may look at a consumer's credit history and previous transactions to see if they're a credit risk in the future.

Association:

Association determines links between different variables. In this situation, you'll look for certain events that are linked with one another; for example, you might discover that when your consumers buy one thing, they frequently buy another, related item. This is commonly used to populate "people also bought" sections on online stores.

Classification :

Classification is an advanced data mining technique that requires you to group together diverse attributes into discernible groups, which you can then use to make additional conclusions or perform a specific job. You might be able to designate individual consumers as "low," "medium," or "high" credit risks based on data about their financial backgrounds and buying history. These classifications might then be used to learn even more about those clients.

Clustering:

Clustering is similar to classification, in that it involves putting together groups of data based on their commonalities. For example, you may group different demographics of your audience into distinct categories

based on their discretionary income or how frequently they purchase at your store.

Outlier detection:
In many circumstances, simply finding the overall pattern will not provide you with a complete picture of your data. You must also be able to spot anomalies, sometimes known as outliers, in your data.

Regression:
Regression is a type of planning and modelling that is used to determine the probability of a particular variable, given the presence of other variables. You may use it, for example, to forecast a price based on other criteria such as availability, consumer demand, and competition. The main goal of regression is to help you figure out the relationship between several variables in a data set.

APPLICATIONS OF DATA MINING

Shopping market analysis:
In the shopping market, there is a big quantity of data, and the user must manage enormous amounts of data using various patterns. To do the study, market basket analysis is a modeling approach.
Market basket analysis is basically a modeling approach that is based on the notion that if you purchase one set of products, you're more likely to purchase another set of items. This strategy may help a retailer understand a buyer's purchasing habits. Using differential analysis, data from different businesses and consumers from different demographic groups may be compared.

Weather Forecasting analysis:
For prediction, weather forecasting systems rely on massive amounts of historical data. Because massive amounts of data are being processed, the appropriate data mining approach must be used.

Stock market analysis:
In the stock market, there is a massive amount of data to be analyzed. As a result, data mining techniques are utilized to model such data in order to do the analysis.

Intrusion detection:
Well, data mining can assist to enhance intrusion detection by focusing on anomaly detection. It assists an analyst in distinguishing between unusual network activity and normal network activity.

Fraud detection:
Traditional techniques of fraud detection are time-consuming and difficult due to the amount of data. Data mining aids in the discovery of relevant patterns and the transformation of data into information.

Surveillance:
Well, video surveillance is utilized practically everywhere in everyday life for security perception. Because we must deal with a huge volume of acquired data, data mining is employed in video surveillance.

Financial banking:
With each new transaction in computerized banking, a massive amount of data is expected to be created. By identifying patterns, causalities, and correlations in corporate data, data mining may help solve business challenges in banking and finance.

CONCLUSION

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. Today increasingly more companies acknowledge the value of this new opportunity and use data mining tools and solutions that help optimizing their operations and increase customer's bottom line.

REFERENCE

[1] S. Agarwal, "Data Mining: Data Mining Concepts and Techniques," 2013 International Conference on Machine Intelligence and Research Advancement, Katra, India, 2013, pp. 203-207, doi: 10.1109/ICMIRA.2013.45.

[2] M. Li, H. Wang and J. Li, "Mining conditional functional dependency rules on big data," in Big Data Mining and Analytics, vol. 3, no. 1, pp. 68-84, March 2020, doi: 10.26599/BDMA.2019.9020019.

[3] A. Marquardt et al., "HealthSCOPE: An Interactive Distributed Data Mining Framework for Scalable Prediction of Healthcare Costs," 2014 IEEE International Conference on Data Mining Workshop, Shenzhen, China, 2014, pp. 1227-1230, doi: 10.1109/ICDMW.2014.45.