International Journal of Innovative Research in Engineering
Volume 3, Issue 5 (September-October 2022), PP: 113-114.
www.theijire.com **ISSN No: 2582-8746**

Data Mining

Swetha A[1], Raksitha Sree S[2], Kowsika K[3], Kalai Selvi K[4]
[1,2,3,4] *Computer Science, Sri GVG Visalakshi College for Women, Bharathiar University, tamilnadu, India.*

How to cite this paper:
Swetha A[1], Raksitha Sree S[2], Kowsika K[3], Kalai Selvi K[4], "Data Mining", IJIRE-V3I05-113-114.

Abstract: *Advancement in technologies has led to the emergence of various new fields. Different fields like science, Engineering, Health, Business are generating, accumulating high amounts of data every day. Data mining is a vital area, which manages and extracts required information from these enormous amounts of data. This paper gives a gist of how data mining is utilized in different fields.*

Keyword: *Machine learning, information retrival.*

I. INTRODUCTION

Data mining is the process of analysing enormous amounts of information and datasets, extracting (or "mining") useful intelligence to help organizations solve problems, predict trends, mitigate risks, and find new opportunities. Data mining is like actual mining because, in both cases, the miners are sifting through mountains of material to find valuable resources and elements. Data mining also includes establishing relationships and finding patterns, anomalies, and correlations to tackle issues, creating actionable information in the process. Data mining is sometimes called Knowledge Discovery in Data (KDD).

II.DATA MINING STEPS

- Understand Business
- Understand the data
- Prepare the data
- Model the data
- Evaluate the data
- Deploy the solution

III. DATA MINING TOOLS

As engineers are fond of saying, "Use the right tool for the right job." Here is a selection of tools and techniques that provide data analysts with diverse data mining functionalities.

Artificial Intelligence

AI systems perform analytical functions that mimic human intelligence, such as learning, planning, problem-solving, and reasoning.

Association Rule Learning

This toolset, also called market basket analysis, searches for relationships among data set variables. For example, association rule learning can determine which products are frequently purchased together (e.g., a smart phone and a protective case.

Clustering

This process partitions datasets into a set of meaningful sub-classes, known as clusters. The process helps users understand the natural structure or grouping within the data.

Classification

This technique assigns particular items in a dataset to different target categories or classes. The goal is to develop accurate predictions within the target class for each case in the data.

Data Analytics

The data analytics process enables professionals to evaluate digital information and turn it in to useful business intelligence.

Data Cleansing and Preparation

This technique transforms the data into a form optimal for further analysis and processing. Preparation includes activities such as identifying and removing errors and missing or duplicate data.

Data Warehousing

Data warehousing consists of an extensive collection of business data that businesses use to help them make decisions. Warehousing is a fundamental and necessary component of most large- scale data mining efforts.

Machine Learning

Related to the AI technique mentioned earlier, machine learning is a computer programming technique that employs statistical probabilities to provide computers with the ability to learn without human intervention or being manually programmed.

Regression

The regression technique predicts a range of numeric values in categories such as sales, stock prices, or even temperature. The ranges are based on the information found in a particular dataset.

IV.BENEFITS OF DATA MINING

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 analyse enormous amounts of data quickly
- Data scientists can use the information to detect fraud, build risk models, and improveproduct safety
- It helps data scientists quickly initiate automated predictions of behaviors and trends and discover hidden patterns

V.CONS IN DATA MINING

Data mining techniques are not infallible, so there's always the risk that the information isn't entirely accurate. This obstacle is especially relevant if there's a lack of diversity in the dataset.

Companies can potentially sell the customer data they have gleaned to other businesses and organizations, raising privacy concerns.

Data mining requires large databases, making the process hard to manage.

VI.FINDINGS AND CONCLUSION

At the interface of rising production complexity due to shifting market demands and vast amounts of production data, data mining can be a valid tool to support managing complexity. Most applications of data mining in production management have so far been related to quality management. There are very few applications of DM directly related to production complexity. However, other applications of DM in other fields of production management serve the purpose of managing production complexity very well.

References

1) data mining concepts and techniques- Terry Halpin, Tony Morgan.

2) http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/505_02_00_Kantardzic- Data-Mining_- Concepts-Models-Methods-and-Algorithms-Second-Edition.pdf

3) Data mining application and techniques-M.bharathi

4) Introduction to data mining-second edition

5) Data prepration of data mining- D Pyle