



AN OVERVIEW OF DATA MINING TECHNIQUES AND ITS APPLICATIONS

C.RADHA ¹ | S. PALANIMURUGAN ² | C.MANI ³ | MR.B.MOHANRAJ ⁴

¹ ASSOCIATE PROFESSOR, DEPARTMENT OF MCA, MUTHAYAMMAL ENGINEERING COLLEGE, NAMAKKAL, TAMILNADU, INDIA.

^{2,3,4} STUDENT, I MCA, MUTHAYAMMAL ENGINEERING COLLEGE, NAMAKKAL, TAMILNADU, INDIA.

ABSTRACT:

Data mining is a process which finds useful patterns from large amount of data. The paper discusses few of the data mining techniques, algorithms and some of the organizations which have adapted data mining technology to improve their businesses and found excellent results. In order to determine how data mining techniques (DMT) and their applications have developed, during the past decade, this paper reviews data mining techniques. We'll explore a number of important categories of data mining techniques, such as pattern matching, data visualisation, meta-rule guided mining, generalisation, characterisation, classification, clustering, association, and evolution. We will look at knowledge mining techniques for various database types, such as relational, transactional, object-oriented, spatial, and active databases, as well as global information systems. There will also be a discussion of various research issues and possible uses for data mining. Data mining tools are specialised tools required to analyse and derive meaningful conclusions and inferences from this massive volume of data. An overview of data mining systems and some of their uses is provided in this paper.

KEYWORDS:

DATA MINING, ARCHITECTURE, TECHNIQUES.

PAPER ACCEPTED DATE:

27th April 2024

PAPER PUBLISHED DATE:

1st May 2024

I. INTRODUCTION

Data mining is the process of extracting knowledge or insights from large amounts of data using various statistical and computational techniques. Data mining is the process of searching and analyzing a large batch of raw data to identify patterns and extract useful information.

Data mining can be used in various industries, such as marketing, finance, healthcare, and telecommunications. Data mining can be used in various industries, such as marketing, finance, healthcare, and telecommunications.

DATA MINING TECHNIQUES

- 1) Classification
- 2) Clustering
- 3) Prediction
- 4) Association Rules

A) CLUSTERING

Clustering is an unsupervised learning method. It does not need a training dataset. Clustering is collecting similar objects together to create a group. In some applications, the clustering is called data segmentation. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as pre-processing

approach for attribute subset selection and classification

TYPES OF CLUSTERING METHODS,

- 1) Partitioning Methods
- 2) Hierarchical Agglomerative (divisive) methods
- 3) Density based methods
- 4) Grid-based methods
- 5) Model-based methods

B. CLASSIFICATION

Classification is the most commonly applied data mining technique, which employs a set of pre-classified .examples to develop a model that can classify the population of records at large. Fraud detection and credit-risk applications are particularly well suited to this type of analysis. Classification is a supervised learning method that is considered the most common data mining technique.

TYPES OF CLASSIFICATION MODELS:

1. Classification by decision tree induction
2. Bayesian Classification
3. Neural Networks

4. Support Vector Machines (SVM)

C. PREDICATION

Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables.

PREDICATION TYPES OF REGRESSION METHODS

1. Linear Regression
2. Multivariate Linear Regression
3. Nonlinear Regression
4. Multivariate Nonlinear Regression

D. ASSOCIATION RULE

Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value.

TYPES OF ASSOCIATION RULE

1. Multilevel association rule
2. Multidimensional association rule
3. Quantitative association rule

III. ARCHITECTURE OF DATA MINING

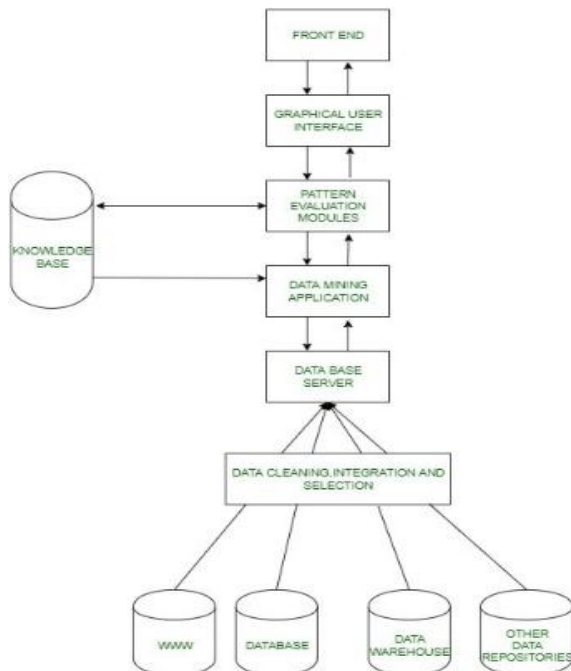


FIG 1: DATA MINING ARCHITECTURE

Data mining architecture is a complex system that involves various components and processes to extract meaningful knowledge from large volumes of data.

DATA SOURCE

- ✓ The data source serves as the origin of data. It can include:
 - ✓ **Databases:** Where structured data is stored.
 - ✓ **Data warehouses:** Repositories containing historical data.
 - ✓ **World Wide Web (WWW):** Unstructured data from websites.
- ✓ The data in these sources may be in the form of plain text, spreadsheets, or other forms of media like photos or videos. WWW is one of the biggest sources of data.

DATABASE SERVER

The database server contains the actual data ready to be processed.

- ✓ The **server** holds the original data ready for processing.
- ✓ It retrieves relevant data based on user requests for data mining.

DATA MINING ENGINE

- ✓ It is one of the core components of the data mining architecture.
- ✓ It contains modules for various data mining tasks:
 - ✓ **Association:** Discovering on the relationships between variables.
 - ✓ **Characterization:** Describing data patterns.
 - ✓ **Classification:** Categorizing data into predefined classes.
 - ✓ **Clustering:** Grouping similar data points.
 - ✓ **Prediction:** Forecasting future trends.
 - ✓ **Time-series analysis:** Analyzing sequential data.

PATTERN EVALUATION MODULE

- ✓ This module assesses patterns discovered by the data mining engine.
- ✓ It uses a **threshold value** to determine interesting patterns.
- ✓ They are responsible for finding interesting patterns in the data and sometimes they also interact with the database servers for producing the result of the user requests.

DATA CLEANING, INTEGRATION, AND SELECTION

- ✓ Before processing, information ought to be cleaned, integrated, and selected.
- ✓ Cleaning includes getting rid of inconsistencies, errors, and duplicates.
- ✓ Integration combines information from diverse reassets right into a unified format.
- ✓ Selection guarantees most effective applicable information is exceeded to the following stage.

IV. APPLICATIONS OF DATA MINING

Data is a set of discrete objective facts about an event or a process that have little use by themselves unless converted into information. We have been collecting numerous data, from simple numerical measurements and text documents to more complex information such as spatial data, multimedia channels, and hypertext documents.

Many of these organizations are combining data mining with such things as statistics, pattern recognition, and other important tools. Data mining can be used to find patterns and connections that would otherwise be difficult to find. This technology is popular with many businesses because it allows them to learn more about their customers and make smart marketing decisions. Here is overview of business problems and solutions found using data mining technology

Data mining provides competitive advantages in the knowledge economy. It does this by providing the maximum knowledge needed to rapidly make valuable business decisions despite the enormous amounts of available data

SCIENTIFIC ANALYSIS

Large amounts of data are produced daily by scientific simulations. This encompasses information obtained from nuclear testing facilities, psychological data, and so forth. Data mining algorithms are able to analyse these kinds of data.

- ✓ **Sequence analysis in bioinformatics:** Identifying patterns in DNA or protein sequences.
- ✓ **Classification of astronomical objects:** Categorizing celestial bodies based on their properties.
- ✓ **Medical decision support:** Assisting doctors in diagnosing diseases and recommending treatments.
- ✓ **Intrusion Detection:** A network intrusion refers to any unauthorized activity on a digital network. Network intrusions often involve stealing valuable network resources.
- ✓ Data mining aids in identifying unauthorized activities on digital networks. It helps detect network intrusions and potential security threats. For example:
 - Detect security violations
 - Misuse Detection
 - Anomaly Detection

HEALTHCARE AND MEDICAL RESEARCH

- ✓ Data mining improves healthcare systems by: Identifying best practices for treatment and cost reduction Analyzing patient data to enhance medical decision-making.
- ✓ A Pharmaceutical sector can examine its new deals force activity and their outcomes to

improve the focusing of high-value physicians

- ✓ Identify successful medical therapies for different illnesses.
- ✓ Characterizes patient behavior to predict office visits.
- **MARKETING AND CUSTOMER RELATIONSHIP MANAGEMENT (CRM)**
- ✓ Organizations use data mining to:
- ✓ **Profile customers:** Understand their preferences and behavior.
- ✓ Perform **market basket analysis:** Identify product associations and recommend related items.
- ✓ Data mining concepts are in use for Sales and marketing to provide better customer service, to improve cross-selling opportunities, to increase direct mail response rates.

EDUCATION

For analyzing the education sector, data mining uses Educational Data Mining (EDM) method.

- Predicting student performance
- Teachers teaching performance
- Curriculum development
- Predicting student placement opportunities

FRAUD DETECTION AND RISK MANAGEMENT

- ✓ Financial institutions use data mining to detect fraudulent transactions and assess risk.
- ✓ By analyzing transaction patterns, they can identify anomalies and prevent fraud.

V. CONCLUSION

- Data mining has importance regarding finding the patterns, forecasting, discovery of knowledge etc., in different business domains. Data mining techniques and algorithms such as classification, clustering etc., helps in finding the patterns to decide upon the future trends in businesses to grow. The user should determine data mining techniques, which technique is used and when. Most of the time the techniques to be used are determined by trial and error. Each technique is problem specific, but in real world data may be continuously changing in dynamic way such as market, finance, spatial and video surveillance. It is not possible to build the perfect model with the dynamic nature of data. But sometimes, it is crucial for decision maker before taking a certain decision. Unpredictable nature of data and chaotic ways data mining techniques are not always predicible.

REFERENCES

1. Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman, 2nd ed.
2. A.S.Aneeshkumar, & C. J. Venkateswaran, 2015, A novel approach for Liver disorder Classification using Data Mining Techniques, Engineering and Scientific International Journal (ESIJ), 2(1): 15-18
3. Dr. Gary Parker, vol 7, 2004, Data Mining: Modules in emerging fields, CD-ROM.
4. Zhao, Kaidi and Liu, Bing, Tirpark, Thomas M. and Weimin, Xiao,"A Visual Data Mining Framework for Convenient Identification of Useful Knowledge", ICDM'05 Proceedings of the Fifth IEEE International Conference on DataMining.vol.-1, no.-1,pp.- 530-537,Dec 2005.
5. S. Alturki, & N. Alturki, 2021, Using Educational Data Mining to Predict Students' Academic Performance for Applying Early Interventions, Journal of Information Technology Education: Innovations in Practice, 20: 121-137.
6. H. Johan, B. Bart and V. Jan, "Using Rule Extraction to Improve the Comprehensibility of Predictive Models".In Open Access publication from Katholieke UniversiteitLeuven, pp.1-56, 2006
7. Giovanni Vigna, Sumit Gwalani, Kavitha Srinivasan, Elizabeth M. BeldingRoyer and Richard A. Kemmerer, " An Intrusion Detection Tool for AODVbased Ad hoc Wireless Networks", IEEE Computer Society Washington, DC, USA ,2004.