IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.67

Volume 5, Issue 4, March 2025

Data Base and Data Mining

Harmi Bhuva, Liza Chowdhury, Nayana Navare, Vanita Zambare, Vaishnavi Navle
First Year Computer Engineering
Guru Gobind Singh Polytechnic, Nashik, Maharashtra, India

Abstract*: KDD, or knowledge discovery in databases, refers to the autonomous exploration and analysis of extensive data resources. This systematic process seeks to identify genuine, valuable, novel, and comprehensible patterns within complex and large datasets. At the heart of this process lies data mining, which entails utilizing algorithms to scrutinize the data and uncover previously unnoticed patterns. The application of theory to evidence facilitates the analysis, prediction, and comprehension of events. The significance of knowledge discovery and data mining is heightened due to the vast accessibility and volume of data available today. It is not unexpected that both academics and practitioners have access to a variety of methodologies due to the rapid development of the field. No single method is universally superior to others. The data encompasses performance evaluation strategies and techniques, illustrates the application of different methods with examples from real-world scenarios and software tools, and seeks to gather an exhaustive list of all relevant techniques developed in the area.*

Keywords: Knowledge Discovery in Database (KDD), Data Repositories, Databases, Mining

I. INTRODUCTION

The following essential features of databases and data management enhance the usefulness and ease of data systems: scalability, efficiency, dependability, simultaneous access management, and complex query languages. In order to maximize the wise use of data repositories, data mining analyzes current data to find patterns.

Knowledge Discovery in Database (KDD)

The automated process of finding hidden patterns, rules, and other intrinsic contents inside large data sets is known as knowledge discovery in databases, or KDD. The goal of data mining (DM) is to find patterns in organized datasets. Knowledge discovery is the process of combining preexisting knowledge and creating new implicit or explicit knowledge from data and information.

While new tacit information is more closely associated with social interaction, new explicit knowledge is primarily discovered through combinations.

Data Repositories

Knowledge Discovery in Databases (KDD) involves automatically identifying patterns in data. A data repository is a storage place for data, which can include large database management systems or collections of databases that collect, manage, and protect valuable datasets for analysis, sharing, and reporting.

Databases

1. Databases are collections of well-structured information that are simple to access, change, and manage. Large volumes of data may be effectively managed and stored with the help of databases.

Important Database Elements:

1. Database management systems, or DBMS, are computer programs that regulate and control database access. Some of the most well-known DBMSs are SQL Server, MySQL, and Oracle.

2. The creation of a conceptual model of the database, comprising entities, properties, and relationships, is the aim of database design.

3. Normalization of databases: Normalization is the process of arranging data to reduce redundancy and enhance data integrity.

DOI: 10.48175/IJARSCT-24047

4. Database security: This involves safeguarding the database against access, use, disclosure, interruption, alteration, and destruction by unauthorized users.Different Database Types

2. 2. Use relational database management systems (RDBMS) to organize data into tables with pre-existing relationships. Examples include MySQL, Oracle, and SQL Server.

3. NoSQL databases: Store data in many formats, such as key-value, document, and graph. Examples include Redis, Cassandra, and Mongo DB.

4. Object-Oriented Databases: These databases use objects to represent real-world items to store information. Examples include Matisse and Gemstone.

3. Time-Series Databases: These databases are designed to store and retrieve vast quantities of time- stamped data. Influx DB and Open TSDB are two examples.

Methods of Data Mining :

Classification: Distributes data points among pre-existing classes or groups.

The second method is clustering, which groups related data components into clusters. Modeling the relationship between variables is known as regression.

Association Rule Mining: Finds trends and correlations between different variables. Decision Trees: Illustrate the process of decision-making.

Categories of Databases

1. Relational databases: Sort information into tables that have connections already made. Examples include SQL Server, Oracle, and MySQL.

2. NoSQL databases: Handle data in different graph, document, and key-value formats. Examples include Redis, Cassandra, and MongoDB.

3. Databases that are object-oriented: Show data as objects that represent actual things. For instance, Matisse and Gemstone.

4. Time-Series Databases: Made to store and retrieve vast amounts of time-stamped data in an effective manner. OpenTSDB and InfluxDB are two examples.

Mining

Procedure of Data Mining

1. Preparing data for mining: This procedure comprises cleaning, transforming, and preparing data.

2. Data Visualization: Data visualization makes use of graphs, charts, and other graphical approaches to understand and communicate data insights.

3. Machine Learning Techniques: Decision trees, clustering, and regression are examples of machine learning techniques used to identify patterns and relationships in data.

4. Text Mining: This technique involves gleaning insights and patterns from unstructured text data.

Applications of Data Mining

1. Business Intelligence: By analyzing consumer behavior, preferences, and trends, data mining is essential to business intelligence.

2. Marketing: Data mining is used in marketing to classify consumers, predict their behavior, and improve marketing tactics.

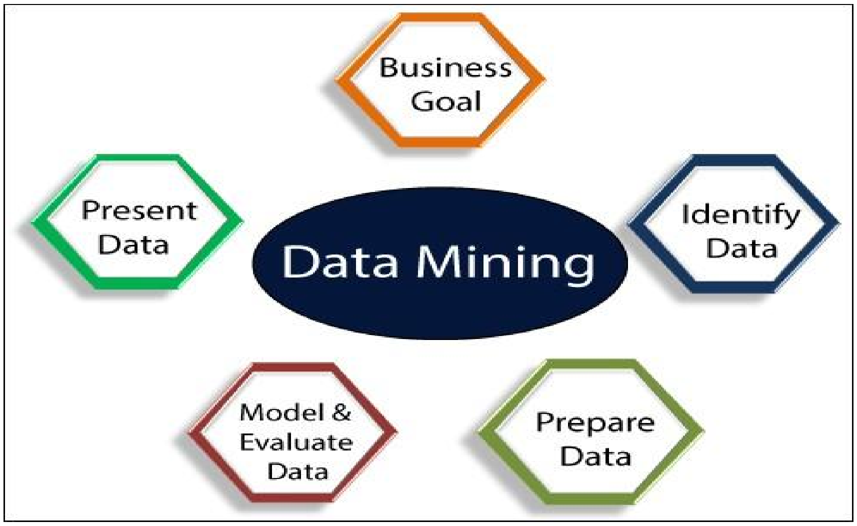
3. Healthcare: By analyzing patient data, predicting disease outcomes, and enhancing treatment strategies, data mining helps the healthcare industry.

4. Finance: Data mining is used in the finance industry to evaluate financial data, forecast market trends, and improve investment portfolios.

Working principle

Principles of Database Functioning

1. Data Acquisition: Data is collected from various sources, such as sensors, user input, and other databases.
2. Data Preservation: A database management system (DBMS) stores the collected data and organizes it into tables, indexes, and other structures.
3. Data Access: The DBMS retrieves the relevant data from storage in response to a user or application request.
4. Data Adjustment: By adding, editing, or deleting records, users can alter the data thanks to the DBMS.
5. Data Inquiry: To obtain specific information, users can use SQL (Structured Query Language) or other query languages to access the database.



The Working Principle of data Mining

1. Data Collection: A range of sources, such as files, databases, and online repositories, are used to collect information.
2. Data Cleaning: The collected data is cleared of mistakes, inconsistencies, and extraneous information.
3. Data Transformation: The cleaned data is converted into an analysis-ready format.
4. Pattern Discovery: The changed data is subjected to data mining techniques such as decision trees, clustering, or regression in order to identify patterns, relationships, and insights.
5. Pattern Evaluation: The reliability, usefulness, and validity of the found patterns are evaluated.
6. Knowledge Representation: Some examples of how the evaluated patterns are usefully represented include reports, infographics, or predictive models.
7. Deployment: Information and insights from data mining are applied to improve operations, inform corporate strategy, and strengthen decision-making.

Methods of Data Mining

1. Classification: The process of allocating data points to pre-existing classes or groupings.
2. Clustering: Putting together groups of relevant data items.
3. Regression: Modeling the relationship between variables.
4. Association Rule Mining: Finding correlations and patterns among variables.
5. Decision Trees: Putting the process of making decisions into visual form.
6. Neural Networks: Artificial neural networks are used to model intricate interactions.
7. Text mining is the process of gleaning patterns and insights from unstructured text data.

Tools for Data Mining

1. R: A popular computer language for statistical calculation and data visualization.
2. Python: A popular programming language for data science and machine learning.

DOI: 10.48175/IJARSCT-24047

3. SQL: A popular language for managing relational databases.
4. Tableau: An interactive tool for creating data visualization dashboards.
5. SAS: A group of tools for data manipulation, statistical analysis, and visualization.
6. Apache Spark: An open-source engine for processing data that can be used for extensive analytics.
7. Microsoft Power BI: An analytics tool for corporate intelligence and data visualization.

REFERENCES

[1]. https://www.researchgate.net/publication/49616224_Data_mining_techniques_and_application
[2]. https://engineering.case.edu/computer-and-data-sciences/research/databases-and-data-mining -
[3]. text=Database%20and%20data%20management%20provide,intelligent%20use%20of%20data%20repositories
[4]. .https://images.app.goo.gl/jj1htMXS6fKKBVTg9