International Journal of Engineering Research & Management Technology
(Peer-Reviewed, Open Access, Fully Refereed International Journal) ISSN: 2348-4039
Volume 11, Issue-1 January-February- 2024 Impact Factor: 7.09
Email: editor@ijermt.org www.ijermt.org

EVALUATING THE BEST ALGORITHMS AND TECHNIQUES FOR MODERN APPLICATIONS OF DATA MINING

Aditi Goel
Research Scholar, School of Technology and Computer Science
The Glocal University, Mirzapur Pole, Saharanpur (U.P) India.

Dr. Geetu Soni
Research Supervisor, School of Technology and Computer Science
The Glocal University, Mirzapur Pole, Saharanpur (U.P) India.

ABSTRACT
Data mining is a vital area within computer science that leverages statistical methods to uncover patterns in databases. Its primary objective is to extract valuable and actionable information from data and transform it into an understandable structure for future use. Various techniques are available to facilitate successful data mining processes. This paper explores a comparative analysis of data mining techniques and algorithms, offering insights into their functionality and applications. Additionally, it highlights cases where corporations have upgraded their data mining technologies, leading to enhanced profits and impressive outcomes. This study aims to deepen understanding of data mining's evolving landscape.

Keywords: Data mining techniques, data mining technology, computer science, statistics, algorithms.

1. INTRODUCTION
The advancement of Information Technology has led to an abundance of data across various domains. Research in information science and technology has established methodologies for the effective storage and utilization of this critical information, facilitating informed decision-making. Data mining serves as a technique for extracting valuable insights and patterns from large datasets. This process, often referred to as information retrieval or data analysis, involves the systematic examination of extensive information to derive meaningful data. The primary objective of data mining is to uncover previously unrecognized patterns that can be leveraged to enhance business decision-making. Although the algorithms employed in data mining have been in existence for over a decade, they have evolved into robust, reliable, and comprehensible tools that surpass traditional methods in sustainability and effectiveness.

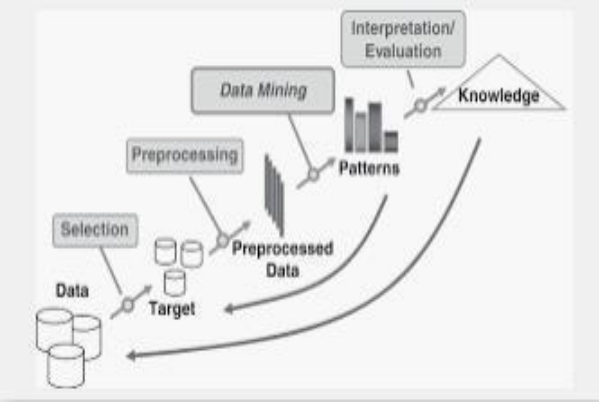

Fig.-1: Knowledge discovery process

Knowledge Discovery in Databases (KDD) refers to the nontrivial Withdrawal of implied, recently unknown and likely useful details from statistics in databases. While data mining and KDD are often handled as synonyms, data mining is genuinely part of the knowledge discovery process.

Three steps involved in KDD are:

- Exploration:

In the first step of statistics exploration records are cleaned and modified into another form, and essential variables and then nature of statistics based on the trouble are determined.

- Pattern Identification:

Once information is explored, refined and described for the particular variables the second step is to structure pattern identification. Identify and select the patterns which make the excellent prediction.

- Deployment:

The Patterns are deployed for preferred outcome.

2. LITERATURE REVIEW

This paper provides information on other data mining algorithms and algorithms for comparative analysis. In this paper we have identified the various types of data that may be mine in this process. We have also reviewed the documentation of specific data mining techniques such as integration, organizational rules etc.

Han et al (2021) - Data mining can be defined as a means of exploring archives and obtaining vague but useful information.

Sumathi et al (2016) - Data mining has the potential to discover hidden relationships and expose unknown patterns and styles by digging into big data.

Hui et al (2016) -The purpose of data mining can be categorized according to the work performed: integration, fragmentation, integration, and retrieval.

3.RELATED WORK

There are many types of data available worldwide. In this paper, we have disclosed various types of data that could be mine with the help of this process. Basically, data mining is not just one type of media or

data. Data mining should be accompanied by any type of data storage. However, algorithms and strategies can be flexible when used with certain types of information. Indeed, the challenges posed by specific types of information vary widely. Data mining is primarily used for information, such as relational information, related data and targeted information, and for trade information, informal and well-structured repositories such as the World Wide Web, high profile information such as location data, multimedia information, timeline information and textual information even flat files. Here are some examples in more detail:

• Flat Files: Flat files are the most common source of data mining algorithm for data mining. It is a basic level of data file in a binary format with a known pattern of data mining the algorithm to be used. Details can be transactions, timeline data, and science rating.

• Relationship database: The relationship database includes a set of tables that contain the values of the trademark, or the values of the symbols from the business relationship. Tables have columns and rows, where columns indicate attributes and rows indicating multiples.

• Database: Database, is a database that is taken from a lot of data and is calculated to be fully utilized under the same integrated system. The data repository provides an opportunity to analyze records from specific sources under the same framework.

4. METHODOLOGY

There are many data mining programs that are accessible or upgraded. Some are special structures that are bound to a given source of information or are limited to limited data mining operations. The structure of a data mine can be categorized according to various conditions namely:

• **Classification about the type of data source mined:** This section classifies data mining systems depending on the type of records managed such as location data, multimedia data, chronological data, textdata, World Wide Web, etc.

• **Classification according to the data model drawn on:** This section classifies data mining systems according to the relevant reality model such as the relationship database, object-oriented database, database, transactions, etc.

Classification according to knowledge discovered: This classification classifies data mining systems based on data types or data mining operations, such as demographics, discrimination, mergers, segregation, mergers, etc.

• **Classification according to mining techniques used:** The data mining structure selects and provides specific strategies. This classification separates data mining systems according to the data analysis method used such as machine learning, neural networks, genetic algorithms, statistics, observations, database or data storage, etc.

There are many tools available for data search. However, we will introduce the most important data mining tools. Also, an analysis of those tools used in recent years.

Data Mining Tools	2016	2017	2023	2019
R	38.5%	46.9%	49%	52%
Rapid Miner	44.2%	31.5%	32.6%	32.8%
SQL	25.3%	30.9%	35.5%	34.9%

Python	19.5%	30.3%	45.8%	52.6%
Excel	25.8%	22.9%	33.6%	28.1%
KNIME	15.0%	20.0%	18.0%	19.1%
Hadoop	12.7%	18.4%	22.1%	15.0%
Tableau	9.1%	12.4%	18.5%	19.4%
SAS	10.9%	11.3%	5.6%	9.12%
Spark	2.6%	11.3%	21.6%	22.7%

Table3: Tools of data mining

III. STANDARD COMPARISON OF CLASSIFIERS ALGORITHMS

The data mining algorithm is widely used in Artificial Intelligence and Machine Learning. They are too many algorithms are available in data mines namely:

Table-1: Advantage and disadvantages of classifier

Classifier	Method	Advantages	Disadvantages
The Support Vector Machine	These are supervised learning models with related learning algorithms that examines data.	1. Highly Accurate. 2. Able to model complicated nonlinear decision boundaries	1. High algorithmic complexity and extensive memory. 2. The choice of the kernel is difficult.
K Nearest Neighbor	An object is classified by a majority vote of its neighbors, most common amongst its k nearest neighbors.	1. Analytically tractable. 2. Simple in implementation. 3. Uses local information, which can submit highly adaptive behavior.	1. Large storage requirements. 2. Highly susceptible to the curse of dimensionality. 3. Slow in classifying test tuples.
Artificial Neural Network	It changes its structure based on its external or internal information.	1. Requires less formal and statistical training. 2. High tolerance to noisy data.	1. Black box nature. 2. Proneness to over fitting.

Bayesian Method	Algorithm attempts to estimate the conditional probabilities of classes given	1. Naïve Bayesian classifier simplifies the computations. 2. Exhibit high accuracy and speed.	1. The assumptions made in class conditional independence. 2. Lack of available data.

5. COMPARISON OF DATA MINING METHODS

Classification:

Separation is a frequently used method of data mining, basically it helps to set up a few pre-programmed examples to update a model that can differentiate the amount of data in general. This method always uses algorithms for neural network configuration. The method of dividing mathematics involves reading and division.

Types of category models:

Bayesian classification

• Neural Networks

Vector support machine (SVM)

• Separation by tree planting.

• Separation based on organization.

Clustering:

Integration can be defined as the identification of categories of comparable objects. By using confluence methods, we can similarly find compact and small spaces in an object's space and we can obtain a universal distribution sample and a correlation between mathematical symbols. Integration can be used as a pre-processing strategy for the subset of adjectives and adjectives.

Types of meeting methods:

• Divide Methods

• Roads designed for congestion

• Grid-based approaches

• Model-based approaches

Predicting:

The return method can be configured to predict. Reversal analysis can be used to maintain the relationship between one or more neutral variables and systematic variables. In the data mineral the random variable is a pre-perceived attribute and the response variable is what we choose to guess.

Types of retrieval methods:

• Corresponding Modifications

• Reduction of Multivariate Linear lines

• Indirect postponement

• Multivariate Nonlinear Regression

Association rule:

Integration and adjustment usually find the findings of setting a common object between large data sets. This type of acquisition helps companies make informed decisions, such as catalog design, opposite marketing and customer behavior analysis. Association Rule algorithms require the ability to produce guidelines with less than one confidence rating.

Types of organization law:

• Multilevel organization law

• Multi-sectoral law

• Measurement organization law

Neural Networks:

The Neural Network is a collection of plug-ins in and out of gadgets and all communications are weighty as well. Neural networks have great potential for finding meaning from complex data and can be used to retrieve patterns and find patterns that are too complex to be seen by humans or other computer systems.

Types of neural networks: Back Propagation

Table-2: Comparison of techniques

Techniques	Average Accuracy
Classification	83.10%
Clustering	82.07%
Predication	82.76%
Association rule	74.72%
Neural Networks	82.85%

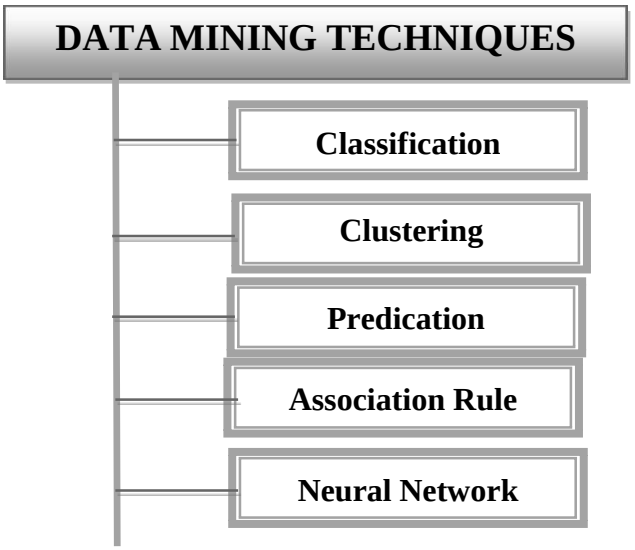

Fig.-2: Data mining techniques

6. CONCLUSION

Data mining plays a critical role in identifying patterns, making predictions, and extracting valuable information across various business domains. It has become an indispensable tool for analyzing and understanding large datasets, making it highly relevant in today's data-driven industries. Techniques such as segmentation, consolidation, and others enable businesses to uncover meaningful patterns, helping them anticipate future trends and make informed decisions.The applications of data mining are vast and span numerous sectors within the data processing industry. As a result, data mining is considered a fundamental component of modern database systems and information management frameworks. It holds immense promise for advancing various fields within Information Technology, driving innovation and efficiency.This paper delves into the comparative analysis of data mining techniques and algorithms, exploring their functionality, advantages, and limitations. Furthermore, it examines how data mining systems can be categorized into distinct processes, highlighting their contributions to different stages of data analysis. By providing insights into these techniques and systems, this study aims to enhance understanding of data mining's impact on business growth and technological development, reinforcing its significance in the evolving landscape of information technology.

7. REFERENCE

[1] Berina Alic, Lejla Gurbeta, Almir Badnjevic, Review Of Machine Learning Activities (June 2017).

[2] Smit Garg, Arvind K Sharma, Comparative Analysis Of Data Details (July 2023).

[3] Keshav Singh Rawat, Comparative Comprehension Of Data Conduct, Algorithm Materials AndMachinery For Actual Data Analysis Reading (July 2017).

[4] Mr. Nilesh Kumar Dokania And Ms. Navneet Kaur, A Comprehensive Study Of Different Skills OfData Information (May 2023).

[5] Mrs. Bharati M. Ramageri, Diy Process And Applications (7 April 2020).

[6] Jiawei Han And Micheline Kamber Jian Pei, Analysis Of Thoughts And Diy Skills (Feb 2016).

[7] M.S Chen, J.Han, And P.S. Yu, A Comparative Study Of Details Of Details And Details (July 2017).

[8] Mr. Nilesh Kumar Dokania And Ms. Navneet Kaur, A Comprehensive Study Of Different Skills OfData Information (May 2023).