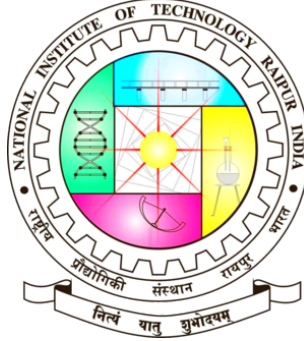


National Institute Of Technology, Raipur



DATA MINING

Dubi Sao, 19111025 Artificial Intelligence ,
VI Semester, Biomedical Engineering Department,

Guided by :
Dr. Saurabh Gupta

Abstract

This paper deals with the study of Data Mining, its importance, advantages and disadvantages and techniques used in it. Data mining is one of the most helpful approaches for extracting valuable information from large sets of data by enterprises, researchers, and individuals. Also known as Knowledge Discovery in Databases (KDD), it finds many applications in various fields.

TABLE OF CONTENT

Contents

1	Introduction	3
2	Importance of Data Mining	4
3	Types of Data Mining	4
4	Data Mining Techniques	5
5	Advantages and Disadvantages of Data Mining	6
6	Applications of Data Mining	7
7	Data Mining in Healthcare	9
8	Conclusion	11
9	References	11

1 Introduction

The process of detecting anomalies, patterns, and correlations within massive databases in order to forecast future outcomes is known as data mining. This is accomplished by combining three fields that are intertwined: statistics, artificial intelligence, and machine learning. Data mining is an automated technique that involves looking for patterns in vast information that humans might miss.

Data mining approaches, for example, are used in weather forecasting. Weather forecasting involves analysing large amounts of past data to find trends and forecast future weather conditions based on the time of year, climate, and other factors.

As a result of this analysis, algorithms or models are developed that collect and analyse data in order to anticipate outcomes with increasing accuracy.



2 Importance of Data Mining

Data mining is an important part of every organization's analytics programme. The data it generates can be used in BI and advanced analytics programmes that analyse historical data, as well as real-time analytics systems that look at data as it's being created or collected.

Effective data mining benefits in several elements of business strategy development and operations management. Marketing, advertising, sales, and customer service are examples of client-facing functions, as well as manufacturing, supply chain management, finance, and human resources. Fraud detection, risk management, cybersecurity planning, and a variety of other key corporate use cases are all aided by data mining. Healthcare, government, scientific research, mathematics, athletics, and other fields all benefit from it.

3 Types of Data Mining

The main types of data that can be used for data mining are as follows:

1. Relational Database:

A relational database is a collection of several data sets that are formally organised by tables, records, and columns that may be accessed in a variety of ways without knowing the database tables. Tables let people find and share information, making data search, reporting, and organisation easier.

2.Data Warehouses:

A Data Warehouse is a piece of software that gathers data from multiple sources within a company in order to deliver useful business insights. The massive volume of data comes from a variety of sources, including Marketing and Finance. The retrieved data is used for analytical purposes and assists a corporate organisation in making decisions. Rather than transaction processing, the data warehouse is intended for data analysis.

3.Data Repositories:

The Data Repository is a broad term for a data storage location. Many IT professionals, on the other hand, use the phrase to refer to a certain type of arrangement within an IT organisation. For example, a collection of databases in which a company has stored numerous types of data.

4.Object-Relational Database:

An object-relational model is a hybrid of an object-oriented database model with a relational database model. It supports Classes, Objects, and Inheritance, among other things. One of the main goals of the object-relational data model is to bridge the gap between relational databases and the object-oriented model techniques common in many programming languages, such as C++, Java, and C.

4 Data Mining Techniques

To turn enormous amounts of data into meaningful information, data mining employs a variety of algorithms and methodologies. Here are a few of the most popular:

1.Association Rules: The term "association rule" refers to a rule-based method for determining associations between variables in a dataset. linkages between different items, typically employs these methodologies. Businesses may develop stronger cross-selling strategies and recommendation engines by understanding their customers' consumption habits.

2. Neural Networks: Neural networks, which are mostly used for deep learning algorithms, analyse training data by simulating the interconnection of the human brain through layers of nodes. Inputs, weights, a bias (or threshold), and an output make up each node. If the output value reaches a certain threshold, the node "fires" or "activates," sending data to the network's next

layer. Through supervised learning, neural networks learn this mapping function, then alter it based on the loss function using gradient descent. We can be sure in the model's accuracy to produce the correct answer when the cost function is at or near zero.

3. Clustering: It is the partitioning of data into groups of related items. The data is described by a few clusters, which sacrifices some details but improves the overall quality. It uses clusters to model data. Clustering is historically anchored in statistics, mathematics, and numerical analysis, as evidenced by data modelling. Clusters are related to hidden patterns in machine learning, the search for clusters is unsupervised learning, and the framework that follows is a data idea. Clustering is quite useful in data mining applications from a practical standpoint. Scientific data exploration, text mining, information retrieval, spatial database applications, CRM, Web analysis, computational biology, medical diagnostics, and a variety of other applications are just a few examples.

4. Regression: Regression analysis is a data mining technique for identifying and analysing the relationship between variables when another factor is present. It is used to specify the likelihood of a particular variable. Regression is essentially a planning and modelling technique. We might use it, for example, to forecast specific costs based on other criteria like supply, consumer demand, and competition. It primarily provides the precise link between two or more variables in a data set.

5 Advantages and Disadvantages of Data Mining

Advantages:

1. Organizations can collect knowledge-based data using the Data Mining technique.
2. Data mining enables businesses to make profitable changes to their operations and production.
3. Data mining is a cost-effective alternative to conventional statistical data

applications.

- 4.Data mining aids an organization's decision-making process.
- 5.It allows for the automatic finding of hidden patterns as well as trend and behaviour prediction.
- 6.It is possible to introduce it into both new and existing systems.
- 7.It's a rapid technique that allows new users to examine large amounts of data in a short amount of time.

Disadvantages:

- 1.There is a chance that businesses will sell valuable customer data to other businesses for a profit. According to the allegation, American Express has sold client credit card purchases to other businesses.
- 2.Many data mining analytics software programmes are difficult to use and require advanced training.
- 3.Because of the various algorithms utilised in their development, different data mining devices operate in different ways. As a result, deciding on the best data mining tools is a difficult undertaking.
- 4.Because data mining techniques are not exact, they may have serious repercussions in certain circumstances.

6 Applications of Data Mining

Retail, communication, financial, and marketing companies employ data mining to assess price, consumer preferences, product placement, and the influence on sales, customer happiness, and corporate profitability. A shop can use point-of-sale records of consumer purchases to generate items and promotions that assist the company attract customers through data mining.



1.Data mining in Education:

Education data mining is a relatively recent field focused with developing ways for extracting knowledge from data generated in educational settings. Students' future learning behaviour is confirmed, the influence of educational support is studied, and learning science is promoted as EDM aims. Data mining can be used by an organisation to make precise judgments and forecast student performance. The institution can now focus on what to teach and how to teach as a result of the findings.

2.Data Mining in CRM (Customer Relationship Management):

It's all about acquiring and retaining customers, as well as improving customer loyalty and implementing customer-centric tactics. A company organisation must collect and evaluate data in order to have a good relationship with its customers. The acquired data can be used for analytics using data mining technology.

3. Data Mining in Fraud Detection:

Traditional fraud detection approaches are time-consuming and complex. Data mining identifies important patterns and transforms data into knowledge. The data of all users should be protected by an ideal fraud detection system. A collection of sample records is used in supervised methods, and these data are classed as fraudulent or non-fraudulent. Using this information, a model is created, and the technique is used to determine whether the document is genuine or not.

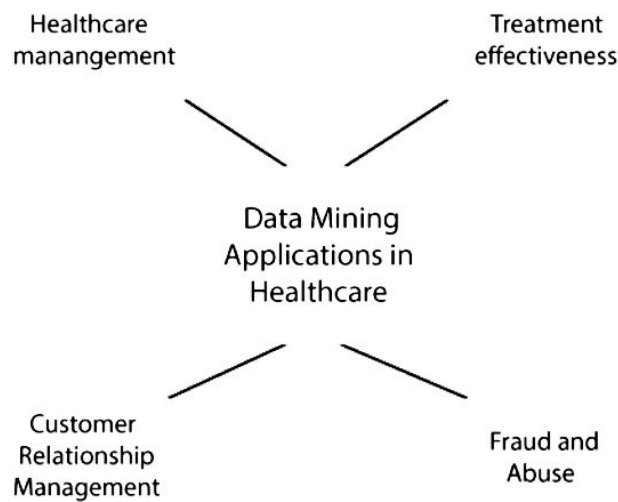
4.Data Mining Financial Banking:

With each new transaction, the financial system's digitalization is expected to generate a massive amount of data. The data mining technique can assist bankers in resolving business-related problems in banking and finance by identifying trends, casualties, and correlations in business information and market costs that are not immediately apparent to managers or executives because the data volume is too large or experts are producing data too quickly on the screen. These data could help the manager better target, acquire, retain, segment, and keep a profitable customer.

7 Data Mining in Healthcare

Due to the exponential growth in the quantity of electronic health records, data mining has enormous potential for healthcare services. Doctors and physicians used to keep patient information on paper, which was difficult to keep track of. Digitalization and the development of new techniques reduce human effort and facilitate data analysis. For example, the computer accurately stores a large volume of patient data, improving the overall data management system's quality. Data mining can aid the healthcare business in the identification and prevention of fraud and abuse, customer relationship management, effective patient care, and best practises, as well as the provision of cheap healthcare services. Conventional approaches cannot process and interpret the massive amounts of data created by healthcare transac-

tions because they are too complex and voluminous. Data mining provides the framework and strategies for transforming these data into information that can be used to make data-driven decisions.



|

1. Treatment effectiveness: Data mining software can be used to evaluate the efficacy of medical treatments. By comparing and contrasting causes, symptoms, and treatment regimens, data mining can provide analysis of which course of action is most beneficial.

2. Healthcare management: Data mining tools can help identify and track chronic illness conditions and incentive care unit patients, reduce hospital admissions, and improve healthcare administration. Data mining is a technique for analysing large data sets and statistics in order to look for patterns that could indicate a bio-terrorist attack.

3. Customer relationship management: Customer and management interac-

tions are critical for any firm to achieve its objectives. Customer relationship management (CRM) is the most common way of managing interactions between commercial enterprises and their clients, usually in the retail and banking sectors. In the same way, it's critical in the healthcare field. Call centres, billing departments, and ambulatory care settings are all places where customers can interact,.

4. Fraud and abuse: Inappropriate or incorrect prescriptions, as well as fraudulent insurance and medical claims, might be the target of data mining fraud and abuse applications.

8 Conclusion

We have successfully discussed the topic Data Mining, its types, importance, techniques used in it as well as its advantages and disadvantages. Furthermore, we have gained insights into its applications and contribution in various fields like retail, consumer, marketing as well as in the healthcare industry. Although data mining has a lot of potential, it also has a lot of drawbacks when it comes to implementation. Various obstacles may arise as a result of performance, data, approaches, and techniques, among other things. When the obstacles or problems are accurately identified and adequately resolved, the data mining process becomes effective.

9 References

1. Data Mining Tutorial - Javatpoint
2. What is Data Mining - TechTarget
3. Mohammed J. Zaki and Wagner Meira Jr. 2014. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, USA.