

## DATA MINING AND SOCIAL MEDIA MINING

### 1. What is data mining?

- According to Sun, data mining is the act of finding patterns in big data sets using techniques that lie at the nexus of database systems, statistics, and machine learning. "Its primary goal is to convert unprocessed data into useful information that can be used to make well-informed decisions, optimize processes, and gain a competitive advantage in a variety of fields."
- To extract useful insights from both structured and unstructured data, data mining employs a variety of methods and algorithms. In order to identify consumer patterns, industry trends, and operational efficiency, business analytics relies heavily on the capacity to parse complicated data sets.

The practice of drawing out valuable information from a collection of data, usually from a data warehouse or group of related data sets, is known as data mining. Sorting through massive data sets to find trends, patterns, and links to aid in well-informed planning and decision-making is the main goal of data mining tools, which contain strong statistical, mathematical, and analytics capabilities.

### Data mining key concepts

Data mining, like many other fields, use its own jargon as short cuts to find key ideas. Understanding these ideas is essential to become proficient in data mining and realizing its potential benefits for a company.

- **Data cleansing:** Data cleaning is another name for it. the procedure of fixing mistakes and omissions in data prior to analysis.
- **Model:** the understanding of connections between data, which are frequently represented as rules.
- **Target:** For instance, data mining aims to find high-value clients.
- **Predictors:** the relevant information that points to the goal.
- **Case:** An instance of data, such the details of a specific consumer, is entered into the model to ascertain how it relates to the target. For instance, is there a chance that this client will make additional purchases from you?
- **Market basket analysis:** identifying consumer purchasing habits based on historical purchasing trends, frequently with the use of information gathered from business loyalty programs.
- **Machine learning:** algorithms that search through big data sets for other situations that are same or similar to known cases.

## **2. Different type of data mining:**

- **Predictive Data Mining Analysis:**

Information used to predict future patterns or outcomes based on past and present data is referred to as predictive data. It serves as the foundation for predictive analytics, a subfield of data science that generates predictions using data mining methods, machine learning models, and statistical algorithms.

- Predictive data examples include customer purchase history, which can be used to forecast future purchasing patterns.
- Equipment sensor readings → to predict maintenance requirements.
- Clickstream data from websites helps to predict user drop-off spots or navigation.
- Credit scores and transaction history are used to forecast the likelihood of loan default
- Weather data can be used to predict energy use or agricultural production.

- **Descriptive Data Mining**

The primary goal of descriptive data mining is to summarize and characterize a dataset's key features. Descriptive data mining looks for patterns, trends, and relationships in past data without attempting to anticipate future events, in contrast to predictive data mining.

- **Structured database**

With the use of a query language like SQL (Structured Query Language), a structured database makes it simple to store, retrieve, and manage data by organizing it into predetermined patterns, usually utilizing tables with rows and columns.

- Deals with: Tabular data in rows and columns (e.g., relational databases, Excel sheets)
- Techniques: SQL queries, traditional data mining algorithms
- Use Case: Sales transaction records, inventory databases

- **Unstructured data**

It refers to information that does not have a predefined data model or organized format, making it difficult to store and manage in traditional relational databases. It often includes text, images, videos, audio, emails, social media content, and other types of data that don't fit neatly into rows and columns.

- Deals with: Data not in a structured format, such as text, images, audio, video
- Techniques: Natural Language Processing (NLP), Computer Vision, Audio analysis
- Use Case: Social media analysis, document classification, image recognition

### **3. Different sources of data mining**

Finding valuable patterns or information in massive databases is known as data mining. Depending on the domain, study style, and data availability, several data sources may be used for mining.

Here are **some of the most common sources of data mining**:

#### ➤ **Relational Databases:**

- One sort of organised data that arranges information into one or more tables, each having rows and columns, is called a relational database. Individual records are represented by rows, while fields or attributes inside those records are represented by columns.
- The basic feature of a relational database is the ability to build associations across distinct tables using a common field called a primary key. This facilitates more effective data retrieval and manipulation by enabling data to be connected and queried across various tables.

#### ➤ **Data Warehouse:**

- The gathering of integrated data from various sources that will be used for queries and decision-making is known as a data warehouse.
- Data warehouses come in three varieties: virtual warehouses, data marts, and enterprise data warehouses.
- Data in Data Warehouse can be updated using either the query-driven approach or the update-driven approach.

**Application:** Business decision making, Data mining, etc.

#### ➤ **Transactional Databases:**

- In order to depict transactions in databases, a collection of data is arranged according to time stamps, dates, and other criteria.
- When a transaction is not committed or finished, this kind of database can reverse or roll back its actions.
- incredibly adaptable technology that allows users to change data without affecting any sensitive data.
- adheres to the DBMS's ACID characteristic.

#### ➤ **Spatial Database**

- Keep track of geographic data.
- holds information in the form of lines, polygons, coordinates, topology, etc.  
Use: Global positioning, maps, etc.

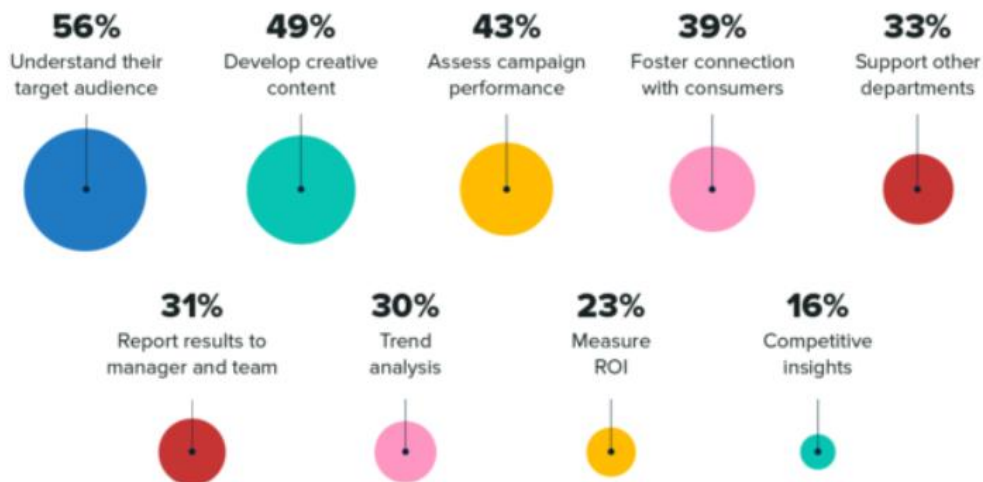
#### ➤ **Time-series Databases**

- Time series databases include user-logged activities and stock market data.

- manages a numeric array that is indexed by time, date, and other factors. Real-time analysis is necessary.
  - Applications include Graphite, Influxdb, eXtremeDB, and others.
- **Streaming Data:**
- This kind of data, which includes log files, social media feeds, and sensor data, is produced continuously.
- **Sensor Data / Iot Devices**
- Information from industrial machinery, smart gadgets, and embedded sensors.
  - Utilized in industries such as manufacturing, healthcare, agriculture, and smart homes.
- **NoSQL Data:**
- This kind of data is kept in a NoSQL database and is accessible via a number of different channels, including document-based, column-based, graph-based, and key-value pair techniques.
- **Cloud Data:**
- Cloud computing environments like AWS, Azure, and GCP are where this kind of data is handled and stored.

#### **4. What is social media mining:**

- The process of gathering, processing, and evaluating the enormous volume of unprocessed data produced on social media sites like Facebook, Instagram, Twitter (or X), LinkedIn, and TikTok is referred to as social media data mining, or social media mining.
- Using a variety of methods and resources, social media data mining aims to identify trends, uncover feelings, and extract pertinent information from user-generated content. Posts, comments, and interactions could all be considered content. Identifying and grouping comparable users, market analysis, sentiment analysis, predictive analytics, and geospatial and temporal studies to find trends and patterns based on time and place are a few noteworthy social media data mining approaches.
- Another crucial element of corporate intelligence is social media mining. It assists in uncovering important information about clients and rivals. It provides crucial information about market conditions. Finding hidden trends and patterns as well as gaining a greater understanding of user behavior, preferences, values, interests, and opinions are the main objectives. When launching new goods, creating marketing plans, or enhancing current procedures or offerings, these insights are utilized to help make well-informed selections.



**Fig: How markets use social data**

**Source:** Sethi, N. (2018, October 23). Data Mining - Working, Characteristics, Types, Applications & Advantages.

Retrieved from <https://electricalfundablog.com/data-mining-working-characteristics-types-applications-advantages/>

## **5. Different technique of data mining :**

**Data scientists and analysts use many different data mining techniques to accomplish their goals. Some of the most common include the following:**

- **Clustering :** Includes finding groups having similar qualities. For instance, clustering is frequently used by marketers to find groups and subgroups within their target audiences. When you are unsure of the commonalities that may be present in your data, clustering can be useful.
- **Classification:** sorts items (or individuals) into categories based on a previously learned model. Classification often comes after clustering (although you can also train a system to classify data based on categories that the data scientist or analyst defines). Clustering identifies the potential groups in an existing data set, and classification puts new data into the appropriate group. Computer vision systems also use classification systems to identify objects in images.
- **Association:** uses a previously learnt model to classify objects (or people) into groups. Clustering is frequently followed by classification, though a system can alternatively be trained to categorize data according to categories defined by the data scientist or analyst. Classification places fresh data into the right group, while clustering finds possible groups in an existing data set. Classification techniques are also used by computer vision systems to recognize items in pictures.
- **Regression:** Predictive analytics frequently uses this more sophisticated statistical method. Developers of mobile apps and social media platforms can use it to boost engagement, predict future sales, and reduce risk. Additionally, regression and

classification can be combined to create a tree model that is applicable in a variety of contexts.

- **Text mining:** examines the frequency of use of specific words. In addition to studying social media posts for marketing objectives or identifying possible employee data leaks, it can be helpful for sentiment or personality analysis.
- **Summarization :** A collection of data is condensed and made easier to grasp through summarization. Summarization can be used, for instance, to compute averages or to make graphs from a given set of data. One of the most well-known and approachable types of data mining is this one.

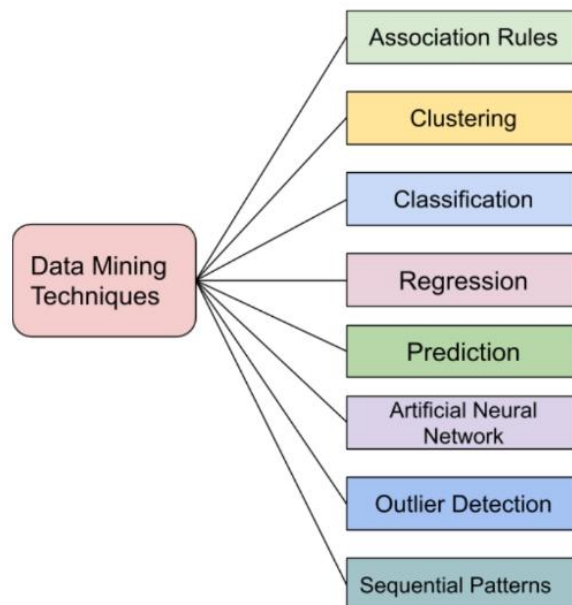
- **Artificial Neural Network:**

A neural network is a group of algorithms that, by simulating the functioning of the human brain, identify underlying relationships in a batch of data. Similar to how the human brain processes information, the artificial neural network (ANN) accomplishes the same.

- **Patterns in Sequence :**

Sequential pattern analysis is a data mining technique that looks at sequential data to find sequential patterns. Finding intriguing subsequences is part of it.

You now have a fundamental understanding of the various data mining approaches. In order to better grasp how these tactics operate, we will outline a few key strategies in the following section.



**Fig: Different type of data mining technique**

**Source:** www.naukri.com. (n.d.). Code 360 by Coding Ninjas. Retrieved from <https://www.naukri.com/code360/library/types-of-data-mining-techniques>

## **Different technique of social media mining:**

In social media, data mining techniques are employed to find patterns and trends in user-generated information. These methods are excellent for examining the information that businesses gather from these sites. Some of the most popular data mining methods for social media were covered in this section.

- **Supervised learning techniques:**

A model is trained to recognize specific characteristics and patterns in the data using supervised learning. This technique is frequently applied in classification domains including sentiment analysis and spam filtering. A label-based dataset is used to train a model. A supervised learning model can be used to categorize consumer comments on social media as either good or negative.

- **Unsupervised learning techniques:**

Learning without the use of training data is known as unsupervised learning. This method is frequently used in social media to identify trends in the data. It can be applied to clustering and anomaly detection. Social media users can be grouped according to their interests or actions using unsupervised learning techniques. Using this approach, the data gathered from users would be analyzed to find trends and commonalities. It might then develop marketing campaigns using this data.

- **Text mining techniques:**

The technique of text mining entails drawing conclusions from user-generated information on social media. It entails finding connections and trends within the text and has a number of uses, including sentiment analysis and entity extraction. Information regarding a text is extracted as part of the sentiment analysis process.

- **Network analysis techniques:**

The study of the connections between different entities, such people and subjects, is referred to as network analysis. It entails displaying the data as a graph or network and analyzing the relationships among its nodes. It is frequently used in social media to examine community members' and influencers' profiles.

Businesses can find influencers and possible clients within a network by using social network analysis. Additionally, it may track these people's activity on the site and cultivate relationships with them. Community discovery is another method that examines the connections between nodes in a network. The social media communities of interest can also be found using this technique.

- **Collaborative Filtering:**

The idea behind collaborative filtering is that it makes recommendations for goods or services based on the tastes and actions of comparable customers. This technique is applied in social media data mining to find pertinent goods and services for consumers.

- **Content-based Filtering:**

Based on the user's viewing history or past preferences, a content-based filtering strategy is used to recommend goods or services. It is employed in social media analysis to locate pertinent material for users according to their search query or history.

- **Predictive Analytics:**

Predictive analysis is a method that forecasts future events based on historical data. It can be applied to social media data mining to predict product demand and boost supply chain effectiveness.



## **REFERENCE:**

- ❖ **Twin, A. (2024, February 23). What is data mining? How it works, benefits, techniques, and examples. Retrieved from <https://www.investopedia.com/terms/d/datamining.asp>**
- ❖ **GeeksforGeeks. (2023, May 6). Different types of data in data mining. Retrieved from <https://www.geeksforgeeks.org/different-types-of-data-in-data-mining/>**
- ❖ **Crockett, E. (2023b, October 17). What is data mining? Types & examples. Retrieved from <https://www.datamation.com/big-data/what-is-data-mining/>**
- ❖ **<https://library.uir.ac.id/home/images/LibraryUIR/ardi/ebook/cambridge-university-press-social-media-mining-an-introduction-2014.pdf>**
- ❖ **SunTec India. (2024, June 20). Social media data mining: Everything you need to know. Retrieved from <https://www.suntecindia.com/blog/social-media-data-mining/>**
- ❖ **What is data mining? | Definition, importance, and types | SAP. (n.d.). Retrieved from <https://www.sap.com/uk/products/data-cloud/hana/what-is-data-mining.html>**
- ❖ **What is data mining? (2024, September 25). Retrieved from <https://business.purdue.edu/master-of-business/online-masters-in-business-administration/posts/what-is-data-mining.php>**
- ❖ **insightsoftware. (2024, October 31). Data Source Types | What Are Data Sources? | insightsoftware. Retrieved from <http://insightsoftware.com/encyclopedia/data-source-types/>**
- ❖ **Different types of social media users | Conversationalists / learners / aggregators / marketers / researchers / lurkers (CLAMRL). (2014, November 18). Retrieved from <https://justadandak.com/different-types-of-social-media-users-connectors-learners-aggregators-marketers-clamrl/>**
- ❖ **<https://neuroquantology.com/open-access/Data+Mining+in+social+media%253A+An+Analysis+of+Techniques+and+Applications+11022/?download=true>**