|  |
| --- |
| RLMCA 301 |
| ASSIGNMENT 1 |
| DATA MINING |

|  |
| --- |
|  |

Submitted by

Adrishya Maria Abraham

Rollno:03

INTMCA S9

Submitted to

Ms. Lisha Varghese

**What is data mining?**

Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis. Data mining techniques and tools enable enterprises to predict future trends and make more-informed business decisions.

Data mining is a key part of data analytics overall and one of the core disciplines in data science, which uses advanced analytics techniques to find useful information in data sets. At a more granular level, data mining is a step in the knowledge discovery in databases (KDD) process, a data science methodology for gathering, processing and analyzing data. Data mining and KDD are sometimes referred to interchangeably, but they're more commonly seen as distinct things.

**Comparison between data mining and big data**

| **Data Mining** | **Big Data** |
| --- | --- |
| It is one of the method in the pipeline of Big Data. | Big Data is a technique to collect, maintain and process the huge information. It explains the data relationship. |
| Data mining is a part of Knowledge Discovery of the Data. It is close view of the data. | It is about extracting the vital and valuable information from huge amount of the data. It is a technique of tracking and discovering of trends of complex data sets. It is a large or overall view of the data. |
| The goal is same as Big Data as it is one of the tool of Big Data. | The goal is to make data more vital and usable i.e. by extracting only important information from the huge data within existing traditional aspects. |
| It is manual as well as automated in nature. | It is only automated as computing huge data is difficult. |
| It only focuses on only one form of data. i.e. structured. | It focuses and works with all form of data i.e. structured, unstructured or semi-structured. |
| It is used to create certain business insights. Data mining is a manager of the mine. | It is mainly used for business purposes and customer satisfaction. Big Data is a mine. |
| It is a sub set of Big Data. i.e. one of the tools. | It is a super set of Data Mining. |
| It is a tool to dig up the vital information from the large data. Data can be large as well as small. | It is more involved with the processes of handling voluminous data. Data can only be large. |

**Web Mining**

Web Mining is the process of Data Mining techniques to automatically discover and extract information from Web documents and services. The main purpose of web mining is discovering useful information from the World-Wide Web and its usage patterns.

**Applications of Web Mining:**

Web mining is the process of discovering patterns, structures, and relationships in web data. It involves using data mining techniques to analyze web data and extract valuable insights. The applications of web mining are wide-ranging and include:

Personalized marketing:

Web mining can be used to analyze customer behavior on websites and social media platforms. This information can be used to create personalized marketing campaigns that target customers based on their interests and preferences.

E-commerce

Web mining can be used to analyze customer behavior on e-commerce websites. This information can be used to improve the user experience and increase sales by recommending products based on customer preferences.

Search engine optimization:

Web mining can be used to analyze search engine queries and search engine results pages (SERPs). This information can be used to improve the visibility of websites in search engine results and increase traffic to the website.

Fraud detection:

Web mining can be used to detect fraudulent activity on websites. This information can be used to prevent financial fraud, identity theft, and other types of online fraud.

Sentiment analysis:

Web mining can be used to analyze social media data and extract sentiment from posts, comments, and reviews. This information can be used to understand customer sentiment towards products and services and make informed business decisions.

Web content analysis:

Web mining can be used to analyze web content and extract valuable information such as keywords, topics, and themes. This information can be used to improve the relevance of web content and optimize search engine rankings.

Customer service:

Web mining can be used to analyze customer service interactions on websites and social media platforms. This information can be used to improve the quality of customer service and identify areas for improvement.

Healthcare:

Web mining can be used to analyze health-related websites and extract valuable information about diseases, treatments, and medications. This information can be used to improve the quality of healthcare and inform medical research.

**Web Data Mining**

Web data mining enables businesses, researchers, and organizations to gain valuable insights from the vast amount of data available on the web. It helps in understanding customer behavior, improving marketing strategies, enhancing user experiences, and making informed business decisions in the digital age.

Here are some common methods and techniques used in web data mining:

Web Crawling:

Web crawlers, also known as web spiders or bots, are used to automatically browse and collect data from websites. These crawlers follow links, visit web pages, and extract relevant information such as text, images, URLs, and metadata. Web crawling is the initial step in web data mining, as it enables the collection of data from a large number of web pages.

Text Mining and Natural Language Processing (NLP):

Text mining techniques are applied to analyze and extract valuable information from textual data obtained from web pages, social media posts, blogs, and other online sources. NLP algorithms are used to process and understand the natural language content, extract entities, perform sentiment analysis, and categorize text into relevant topics or themes.

Web Content Mining:

Web content mining focuses on extracting information and knowledge from the content of web pages. It involves techniques such as information extraction, where specific data fields are identified and extracted from web pages, and web scraping, which involves automated extraction of structured data from HTML pages.

Web Structure Mining:

Web structure mining analyzes the link structure of the web to uncover patterns and relationships between web pages. It involves techniques like link analysis, which examines the links between web pages and determines their importance or relevance based on factors such as link popularity or PageRank algorithm. Web structure mining is used for tasks like web page ranking, recommendation systems, and identifying communities or clusters of related web pages.

Social Media Mining:

Social media platforms generate vast amounts of data in the form of posts, comments, likes, shares, and user profiles. Social media mining involves applying data mining techniques to extract insights from this social data, such as identifying trending topics, analyzing user sentiment, detecting influencers, and understanding social network dynamics.

Web Usage Mining:

Web usage mining focuses on analyzing user behavior and interaction with websites. It involves examining web server logs, clickstream data, and user navigation patterns to understand user preferences, identify popular content, optimize website design, and improve user experience.

Machine Learning and Data Analytics:

Machine learning algorithms and data analytics techniques are applied to web data to discover patterns, predict user behavior, and make data-driven decisions. These techniques include clustering, classification, regression, association rule mining, and anomaly detection.

Web data mining enables businesses, researchers, and organizations to gain valuable insights from the vast amount of data available on the web. It helps in understanding customer behavior, improving marketing strategies, enhancing user experiences, and making informed business decisions in the digital age.