SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
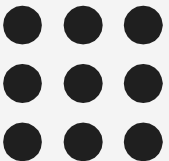Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

Course Name – Data Warehouse & Mining

II Year / IV Semester

Topic – Web Mining

1

Web Mining

- Web mining refers to the process of discovering and extracting useful information from a large amount of data available on the World Wide Web.

- It involves applying various data mining techniques to web data to identify patterns, trends, and relationships.

- Web mining is a multidisciplinary field that combines techniques from data mining, machine learning, artificial intelligence, statistics, and information retrieval.

Types of Web Mining

Web Content Mining

- Process of extracting useful information from web pages, including text, images, and multimedia content.
- This involves techniques such as text mining, natural language processing, and image analysis.
- Web content mining can be used to extract structured and unstructured data from web pages, including

product descriptions, reviews, and user-generated content.

- The extracted information can be used for various purposes, such as sentiment analysis, product recommendation, and opinion mining.

Web Structure Mining

- Web structure mining focuses on analyzing the web structure and the relationships between web pages.
- This includes analyzing links between pages, identifying communities of pages, and detecting patterns in website design.
- Web structure mining techniques are used to improve search engine results, identify authoritative pages, and detect web spam.

Types of Web Mining

Web Usage Mining

- Web usage mining involves analyzing user behavior on the web, including clickstream data, search queries, and other interactions with web pages.
- Web usage mining can help identify user preferences, behavior patterns, and trends.
- This information can be used to personalize content, improve website design, and target advertising.
- Web usage mining can also be used for security purposes, such as detecting fraud and identifying potential security threats.

Process of Web Mining

Data collection - Web data is collected from various sources, including web pages, databases, and APIs.

Data pre-processing - The collected data is pre-processed to remove irrelevant information, such as advertisements and duplicate content.

Data integration - The pre-processed data is integrated and transformed into a structured format for analysis.

Pattern discovery -Web mining techniques are applied to identify patterns, trends, and relationships.

Evaluation -The discovered patterns are evaluated to determine their significance and usefulness.

Visualization -The analysis results are visualized through graphs, charts, and other visualizations.

Difference Between Data Mining and Web Mining

Parameter	Data Mining	Web Mining
Definition	The process of discovering patterns in large datasets	The process of discovering patterns in web data
Data Source	Databases, data warehouses, and other data repositories	Web pages, weblogs, social media, and other web-related data sources
Data Characteristics	Structured, semi-structured, and unstructured data	Mostly unstructured data
Techniques	Clustering, classification, association rules, regression, etc.	Text mining, natural language processing, image analysis, link analysis, etc.
Applications	Marketing, finance, healthcare, etc.	E-commerce, social media, search engines, etc.
Challenges	Data quality, scalability, and privacy concerns	Data heterogeneity, ambiguity, and dynamic nature of the web

THANK YOU