



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai

Accredited by NAAC-UGC with 'A++' Grade (Cycle III) &

Accredited by NBA (B.E - CSE, EEE, ECE, Mech & B.Tech.IT)

COIMBATORE-641 035, TAMIL NADU



COURSE NAME : 19CSE301 INTRODUCTION TO DATA SCIENCE

III YEAR/ V SEMESTER

UNIT – II

Topic: Text Mining

Ms.B.Sumathi

Assistant Professor

Department of Computer Science and Engineering

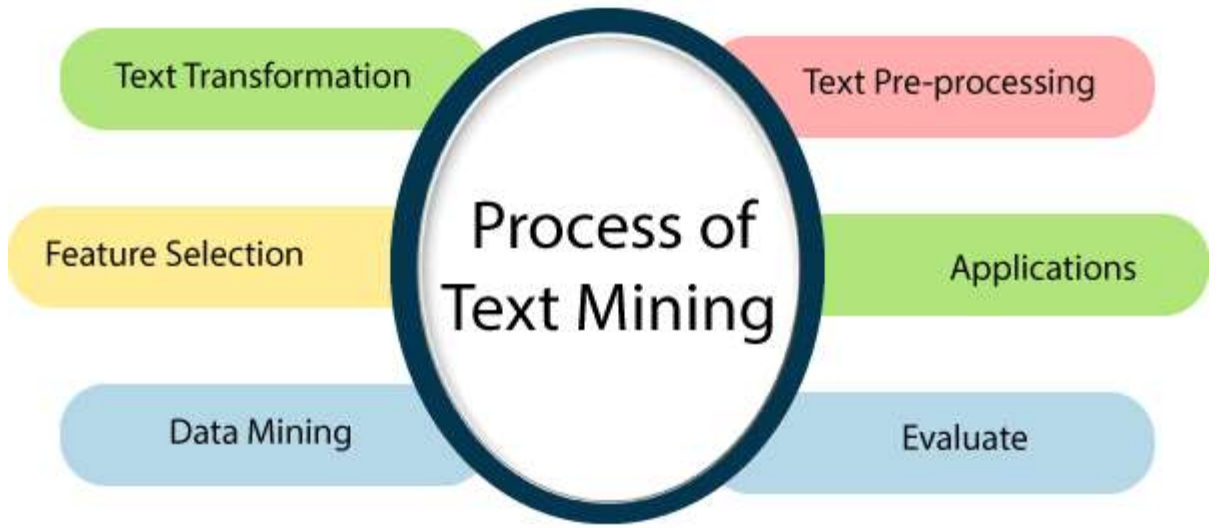
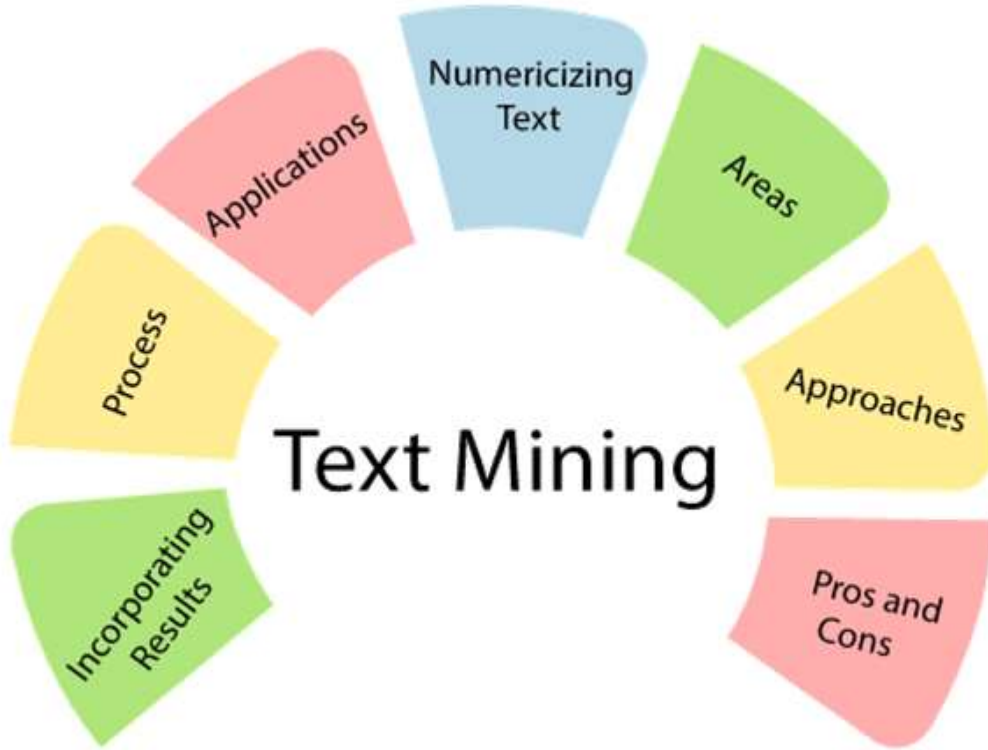


Text Mining

- Text data mining can be described as the process of extracting essential data from standard language text.
- All the data that we generate via text messages, documents, emails, files are written in common language text.



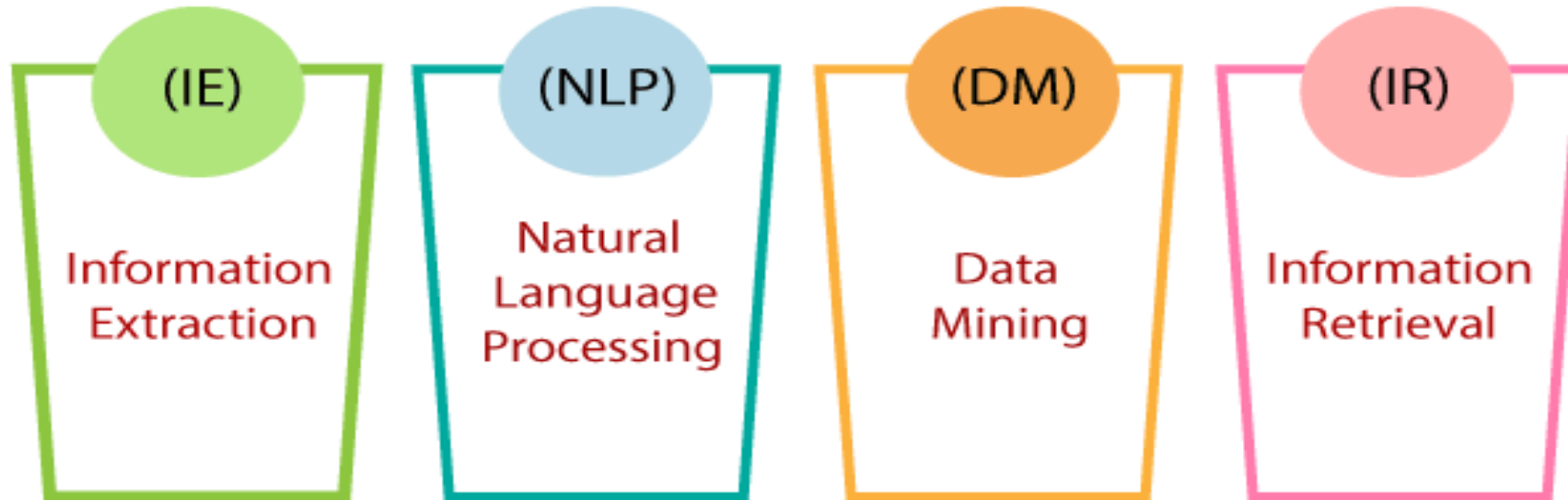
Process's





Area's

Area's of Text Mining





Information Extraction & Retrieval



The automatic extraction of structured data such as entities, entities relationships, and attributes describing entities from an unstructured source is called information extraction.

Information retrieval deals with retrieving useful data from data that is stored in our systems.

Alternately, as an analogy, we can view search engines that happen on websites such as e-commerce sites or any other sites as part of information retrieval



Natural Language Processing

NLP stands for Natural language processing.

- Computer software can understand human language as same as it is spoken.
- NLP is primarily a component of artificial intelligence(AI).
- The development of the NLP application is difficult because computers generally expect humans to "Speak" to them in a programming language that is accurate, clear, and exceptionally structured.
- Human speech is usually not authentic so that it can depend on many complex variables, including slang, social context, and regional dialects.



Data Mining

Data mining refers to the extraction of useful data, hidden patterns from large data sets.

- Data mining tools can predict behaviors and future trends that allow businesses to make a better data-driven decision.
- Data mining tools can be used to resolve many business problems that have traditionally been too time-consuming.



Data Mining

- **Text transformation**

A text transformation is a technique that is used to control the capitalization of the text.

Here the two major way of document representation is given.

- Bag of words
- Vector Space

- **TextPre-processing**

Pre-processing is a significant task and a critical step in Text Mining, Natural Language Processing (NLP), and information retrieval(IR). In the field of text mining, data pre-processing is used for extracting useful information and knowledge from unstructured text data.

Information Retrieval (IR) is a matter of choosing which documents in a collection should be retrieved to fulfill the user's need.



Data Mining

- **Feature selection:**
Feature selection is a significant part of data mining. Feature selection can be defined as the process of reducing the input of processing or finding the essential information sources. The feature selection is also called variable selection.
- **Data Mining:**
Now, in this step, the text mining procedure merges with the conventional process. Classic Data Mining procedures are used in the structural database.
- **Evaluate:**
Afterward, it evaluates the results. Once the result is evaluated, the result abandon.
- **Applications:**
These are the following text mining applications:



Risk Management

- Risk Management is a systematic and logical procedure of analyzing, identifying, treating, and monitoring the risks involved in any action or process in organizations.
- Insufficient risk analysis is usually a leading cause of disappointment. It is particularly true in the financial organizations where adoption of Risk Management Software based on text mining technology can effectively enhance the ability to diminish risk.



Customer Care Service

- Text mining methods, particularly NLP, are finding increasing significance in the field of customer care.
- Organizations are spending in text analytics programming to improve their overall experience by accessing the textual data from different sources such as customer feedback, surveys, customer calls, etc.
- The primary objective of text analysis is to reduce the response time of the organizations and help to address the complaints of the customer rapidly and productively



Business Intelligence

- Companies and business firms have started to use text mining strategies as a major aspect of their business intelligence.
- Besides providing significant insights into customer behavior and trends, text mining strategies also support organizations to analyze the qualities and weaknesses of their opponent's



Social Media Analysis

Social media analysis helps to track the online data, and there are numerous text mining tools designed particularly for performance analysis of social media sites.

- These tools help to monitor and interpret the text generated via the internet from the news, emails, blogs, etc.
- Text mining tools can precisely analyze the total no of posts, followers, and total no of likes of your brand on a social media platform that enables you to understand the response of the individuals who are interacting with your brand and content.



References

- Tom M. Mitchell, “Machine Learning”, McGraw-Hill Education (India) Private Limited, 2013.
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, “An Introduction to Statistical Learning: with Applications in R”, Springer; First Edition 2013.



Thank
You