Semester : **VI**   Subject : **DAV**   Academic Year: 20 **23** 20 **24**.

ROOTS OF TEXT MINING:

Text mining is a component of data mining that deals with unstructured text data.

It involves the use of Natural Language Processing (NLP) techniques to extract useful information and insights from large amount of unstructured text data.

Text mining can be used to extract structured information from unstructured text data such as:

(1) Named Entity Recognition (NER):

Identifying and classifying named entities such as people, organisations, and locations in text data.

(2) Sentiment Analysis:

Identifying and extracting sentiment (eg. positive, negative, neutral) of text data.

Procedure for Analysing Text Mining:

(i) Text summarisation:

To extract its partial content reflects its whole content automatically.

(ii) Text categorisation:

To assign a category to the text among categories predefined by users.

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science
CSE DATA SCIENCE

Semester : **VI**   Subject : **DAV**   Academic Year: 2023- 2024

(iii) Text clustering: To segment text into several clusters, depending on their substantial relevance.

Issues in Text Mining:

Numerous issues happen during the text mining process:

* The efficiency and effectiveness of decision-making.
* The uncertain problem can come at an intermediate stage of text mining. To make the text mining process efficient, different guidelines are developed to ~~no~~ normal the text, in the pre-processing stage.
* Sometimes original message or meaning can be changed due to alteration.
* Many algorithms and techniques support Multi-language text, in text mining. And this creates ambiguity in text mining.
* The use of synonym and antonyms in the document text makes ambiguity in the text mining tools, since text mining tools take both in a similar setting.
* Hence it is difficult to categorise such kinds of text/words.

Advantages of Text mining:

Large amounts of Data: Text mining allows organisations to extract insights from large amounts of unstructural data.

Semester : __VI__   Subject : __DAV__   Academic Year: 2023- 2024.

It includes social media posts, news articles and customer feedback.

Variety of Applications : Text mining has lot many applications such as sentimental analysis, named entity recognition and topic modelling.

Organisation can gain insights from unstructured text data using the above versatile tool.

Improved Decision Making: Text mining can be used to extract insights from unstructured data and that can be used to make data-driven decisions.

Cost-effective : Text mining is a cost-effective way to extract insights from unstructured text data.

Disadvantages of Text Mining:

Complexity : Text mining requires advanced skills in natural language processing and machine learning. This makes it a complex process.

Quality of Data :- The accuracy of insights extracted from text-mining depends on quality of text-data can vary.

High computational cost : Text mining needs high computational resources. And for small organisations it is not feasible.

Limited to text data: Text mining is restricted to extracting insights from unstructured text data and hence cannot be used with other data types.