



Semester : VI

Subject : DAV

Academic Year: 2023 2024

## THE SEVEN PRACTICE AREA OF TEXT ANALYTICS:

Text mining can be divided into seven practice areas, based on unique characteristics of each area. Though distinct, these areas are highly interrelated; a typical text mining project will require techniques from multiple areas. The seven practice areas are as follows:

### 1) Search and Information Retrieval (IR):

Storage and retrieval of text documents, including search engines and keyword search.

### 2) Document clustering:

Grouping and categorizing terms, snippets, paragraphs, or documents, using data mining clustering methods.

### 3) Document classification:

Grouping and categorizing snippets, paragraphs, or documents, using data mining classification methods, based on models trained on labeled examples.

### 4) Web Mining:

Data and text mining on the Internet, with a specific focus on the scale and interconnectedness of the web.



Semester : \_\_\_\_\_

Subject : \_\_\_\_\_

Academic Year: 20 - 20

5) Information Extraction:

Identification and extraction of relevant facts and relationships from unstructured text; the process of making structured data from unstructured and semistructured text.

6) Natural Language Processing: (NLP)

Low-level language processing and understanding tasks (eg. tagging part of speech); often used synonymously with computational linguistics.

7) Concept Extraction:-

Grouping of words and phrases into semantically similar groups.