Unit 1: Introduction to Text Mining

Unstructured Data vs Structured Data, What is Text Mining/Text Analytics?, Various Applications of Text Mining, Text Mining vs Data Mining, Advantages/Benefits of using Text Mining, Limitations/Drawbacks of using Text Mining

Learning outcome: *Gain an overview of text mining and its applications*

Unit 2: Preprocessing - I

Introduction to Text Preprocessing, Importance of Text Preprocessing, Common Text Preprocessing Methods, Why is text standardization required?, NLP Essential Libraries, Text - Preprocessing Python Implementation

Learning outcome: *Describe the basics of preprocessing textual data*

Unit 3: Preprocessing - II

Introduction to Vectorization, What is Vectorization?, Vectorization Techniques, Python Implementation - Introduction to Vectorization

Lab exercise: Implement text data cleaning, preprocessing and vectorizing using Python

Learning outcome**:** *Implement text data cleaning, preprocessing and vectorizing on a given data set*

Unit 4 : Text classification - I

Introduction to Text Classification, Rule Based Systems, Machine Learning Based, Application of Text Classification, Real-world examples for text classification usage

Learning outcome**:** *Discuss the application of text classification*

Unit 5: Text classification - II

Python walkthrough for text classification using various classification algorithms.

Lab Exercise: Implement text classification using Python.

Learning outcome**:** *Implement text classification on a given dataset*

Unit 6: Text Clustering - I

What is clustering in textual data? Application of text clustering: Interpretation, Summarization, Narrows search for similar texts, Similarity measures: Euclidean, Cosine and Manhattan, Edit distance

Learning outcome**:** *Discuss the application of text clustering*

Unit 7: Text Clustering - II

Python walkthrough for text clustering using K-means clustering and hierarchical clustering.

Lab Exercise: Implement text clustering using Python

Learning outcome**:** *Implement text clustering on a given dataset*

Unit 8: Topic Modelling - I

The idea behind topic models, How topic models are learned? Approaches: Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), PCA, ICA, What's the difference between approaches? Part of Speech Tagging.

***Learning outcome*:** *Describe the different approaches to topic modelling*

Unit 9: Topic Modelling - II

Python walkthrough for implementing topic modelling.

Lab Exercise: Implement topic modelling using Python.

***Learning outcome*:** *Implement topic modelling using Python*

Unit 10: Named Entity Recognition - I

What is named entity recognition? How does named entity recognition work? Application of named entity recognition,

***Learning outcome*:** *Describe the basic concepts of named entity recognition*

Unit 11: Named Entity Recognition - II

Python walkthrough for practical implementation of N.E.R. using spacy. **Lab Exercise:** Implement Named Entity Recognition using Python.

***Learning outcome*:** *Implement named entity recognition using Python*

Unit 12: Sentiment Analysis - I

What Is Sentiment Analysis? How Does Sentiment Analysis Work? Why Is Sentiment Analysis Important? Sentiment Analysis Applications.

***Learning outcome*:** *Describe the applications of sentiment analysis*

Unit 13: Sentiment Analysis - II

Python walkthrough for implementing sentiment analysis on a dataset. **Lab Exercise:** Implement sentiment analysis using Python.

***Learning outcome*:** *Implement sentiment analysis using Python*

Unit 14: Text visualization techniques

Why visualize text? Data visualization vs text visualization Word clouds and word tree - word sequences (Advantages and Disadvantages)

***Learning outcome*:** *Describe the basic concepts of text visualisation*

Unit 15: Text mining project

A case study on text mining using Twitter archives.

***Learning outcome*:** *Analyse the application of text mining on a Twitter archive*

Describe the CRISP-DM process.