Module 1

1. Elaborate the process of Text mining and discuss in brief the various algorithms for Text mining. 2
2. Discuss the concept of Named Entity Recognition with example. Explain in detail any one approach to recognize Named entities.
3. Explain the concept of N-gram in text mining. How do unigram, bigram and trigram models differ from each other? Also, discuss the applications of N-gram.
4. Explain the process of Relation Extraction in text mining and discuss its challenges and potential applications.
5. Describe the TF-IDF algorithm and how it is used to represent the importance of words in a document collection.
6. Explain the tokenization technique

Module 2

1. Explain how distance-based clustering algorithms, such as k-means and hierarchical clustering, are applied to text data.
2. What is probabilistic document clustering, and how does it differ from distance-based clustering methods?
3. How do decision tree classifiers work in the context of text classification? What are the advantages and limitations of using decision tree classifiers for text data?
4. Explain the role of Bayesian Networks in text modeling.
5. Explain rule-based classifier.
6. What are proximity-based classifiers, and how are they applied to text classification tasks?
7. Explain how rule-based classifiers are used for text classification.
8. Compare and contrast two distance-based clustering algorithms used in text mining, highlighting their advantages.
9. Describe the process and benefits of word and phrase-based clustering.
10. Discuss advanced clustering techniques such as spectral clustering or DBSCAN in the context of text data.