**Module 1**



1. Elaborate the process of Text mining and discuss in brief the various algorithms for Text mining.



1. Discuss the concept of Named Entity Recognition with example. Explain in detail any one approach to recognize Named entities.



1. 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.



1. Explain the process of Relation Extraction in text mining and discuss its challenges and potential applications.



1. Describe the TF-IDF algorithm and how it is used to represent the importance of words in a document collection.



**Module 2**



1. Explain how distance-based clustering algorithms, such as k-means and hierarchical clustering, are applied to text data.



1. What is probabilistic document clustering, and how does it differ from distance-based clustering methods?



1. 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?



1. Explain the role of Bayesian Networks in text modeling. How can Bayesian Networks be used to represent and infer relationships between different text features?



1. What are proximity-based classifiers, and how are they applied to text classification tasks? Discuss the differences between proximity-based classifiers and other types of classifiers.



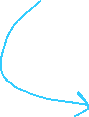
1. Explain how rule-based classifiers are used for text classification. Provide an example of a rule-based classification system and discuss its strengths and weaknesses.



1. Compare and contrast two distance-based clustering algorithms used in text mining, highlighting their advantages.



1. Describe Hidden Markov Models (HMM) and their application in text modeling.



1. Differentiate between two distance-based clustering algorithms used in text mining, highlighting their advantages.
2. What are the common feature selection and transformation methods used in text clustering, and how do they improve clustering performance?



1. Describe the process and benefits of word and phrase-based clustering. How does clustering words or phrases differ from clustering entire documents?



1. Discuss advanced clustering techniques such as spectral clustering or DBSCAN in the context of text data.



**Module 3**

1. Explain the concept of inverted indices in web mining and discuss how compression techniques are applied to optimize them.



1. What are the common techniques used in hiding web spam? Discuss the challenges associated with detecting and combating spam.



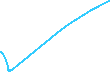
1. How does meta search utilize similarity scores and rank positions to aggregate results from multiple search engines? Provide examples to illustrate.



1. What is Latent Semantic Indexing (LSI) and how does it address the problem of synonymy and polysemy in web search?



1. Discuss the impact of link spamming on search engine algorithms and the quality of search results. How do search engines differentiate between genuine and spammy links?



**Module 4**



1. Explain the process of data pre-processing in web mining. What are the common techniques used to clean and prepare data for analysis?



1. Explain the concept of cluster analysis in web mining. How is it used for visitor segmentation and what are the benefits of segmentation?



1. Describe the concept of data modelling in the context of web mining. How do models help in understanding and predicting user behavior?



1. Discuss the importance of session and visitor analysis in web mining. What insights can be gained from analyzing user sessions and visitor behavior?



1. Explain the analysis of sequential and navigational patterns in web mining. How are sequence mining algorithms applied to discover patterns in user navigation?



1. What is association analysis in web mining? Provide examples of how association rules are derived from web user transactions.



1. What are the main sources and types of data collected in web mining? How does the type of data influence the mining process?



1. Describe the process of classification and prediction based on web user transactions. How are machine learning algorithms used to classify and predict user behavior?



**Module 5**

1. Discuss some of the challenges associated with social network analysis. How do issues like data privacy, scalability, and network dynamics impact the mining of social networks?



1. Describe the different types of social network graphs commonly encountered in social network analysis. Provide examples of each type and explain their significance in understanding social relationships.



1. What is behavior analytics in social media mining? Discuss its importance in understanding user engagement, preferences, and trends on social platforms.



1. Describe classical recommendation algorithms commonly used in social media. Differentiate between collaborative filtering and content-based filtering along with their applications.



1. Explain the concepts of influence and homophily in the context of mining social media. How do these factors influence user behavior and interactions on social platforms?



1. Compare and contrast collaborative filtering and content-based filtering algorithms with examples.



1. What are the key components of an introduction to social network analysis? How do social network graphs facilitate the representation and analysis of social relationships?



1. Outline the challenges involved in recommendation systems for social media. How do factors like data sparsity, user diversity, and evolving preferences affect the effectiveness of recommendation algorithms?



1. Explain the concept of recommendation using social context. How can social relationships and interactions be leveraged to improve the quality and relevance of recommendations?



**Module 6**

1. Discuss abnormal behaviors and group spam detection methods in opinion spam detection. How do these techniques help in identifying and filtering out spammy content?



1. What are the challenges associated with sentiment analysis in multilingual text? How can these challenges be addressed in opinion mining?



1. Compare and contrast supervised and unsupervised approaches to document sentiment classification in opinion mining. Provide examples of each approach.



1. Explain the concept of opinion lexicon expansion in opinion mining. How do dictionary-based and corpus-based methods differ in expanding opinion lexicons?



1. Describe supervised learning techniques used in opinion spam detection. What are the key features considered in identifying opinion spam?



1. Explore the role of domain adaptation in sentiment analysis. How does domain adaptation help in improving the performance of sentiment classifiers across different domains?



1. Describe the process of feature selection in document sentiment classification. What are some common feature selection techniques used in opinion mining?



1. Discuss the limitations of traditional sentiment analysis approaches in handling sarcasm and irony. How can sentiment analysis models be improved to better handle such linguistic nuances?



1. Explain the concept of aspect-based sentiment analysis. How does aspect-based sentiment analysis differ from document-level sentiment classification, and what are its applications in opinion mining?

