DATE: 12-11-2025   TIME: 09.30 a.m. – 12.30 p.m.   MAX.MARKS:20 marks

Answer all the Questions   5 x 20=100 marks

1. a. Briefly explain the key phases in the Natural Language Processing.
   b. Find the probability of test sentence given below if a bigram model is used on the following training sentences. (Note: consider case insensitive corpus data)
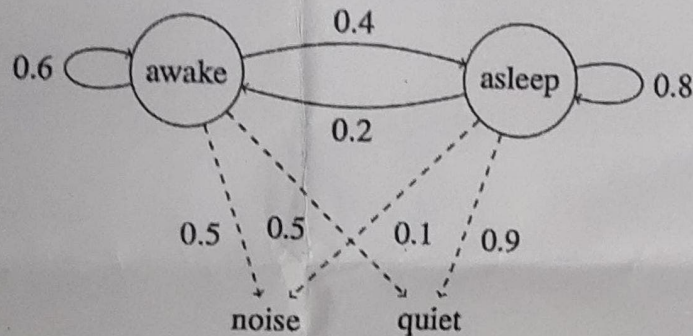   Training data:
   - S1: The Arabian Knights
   - S2: These are the fairy tales of the east
   - S3: The Stories of the Arabian Knights are translated primary languages.

   Test Data: The Arabian Knights are the fairy tales of the east

2. a. Explain Text-to-Speech synthesis(TTS) using concatenative approach.
   b. Find the probability of the observation sequence "noise noise quiet quiet" given the state sequence "awake awake awake asleep" using the Hidden Markov Model. Consider equal initial probability for the states.

3. a. Explain the shallow and dependency parsing with an example.
   b. Perform parsing using top down parsing for the sentence: I saw the man with a telescope.

4. a. Consider there are 1 million documents. Let the document frequency for the words are, auto = 5000, best = 50000, car = 10000, insurance = 1000
   Document 1:   car insurance auto insurance
   Document 2:   best car insurance
   Find the similarity between two documents using cosine similarity. Use TF-IDF for the vector representation of the documents.
   b. Write short notes on feature engineering using Glove and FastText.

5. a. Write short notes on Information Retrieval.
   b. Given the following short movie reviews each labelled with a genre, either comedy or action. Computer the class of the test document using Naïve Bayes Classifier.

	Class	Documents
Training	comedy	fun, couple, love, love
	action	fast, furious, shoot
	comedy	couple, fly, fast, fun, fun
	action	furious, shoot, shoot, fun
	action	fly, fast, shoot, love
Test	?	fast, couple, shoot, fly