Reg No.:_____   Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree 7th semester (S,FE) Exam April 2025 (2019 Scheme)

Course Code: CST473

Course Name: NATURAL LANGUAGE PROCESSING

Max. Marks: 100   **Duration: 3 Hours**

PART A

Answer all questions, each carries 3 marks.   Marks

1	Illustrate Support Vector Machines.	(3)
2	With the help of diagram, explain Machine learning based NLP.	(3)
3	List the shortcomings of one-hot encoding approach.	(3)
4	Differentiate Bag of words with Bag of n-grams.	(3)
5	Consider a sentiment analysis domain with the two classes positive (+) and negative (-), with the following miniature training and test documents simplified from actual movie reviews.	(3)

	Category	Documents
Training	-	just plain boring
	-	entirely predictable and lacks energy
	-	no surprises and very few laughs
	+	very powerful
	+	the most fun film of the summer
Test	?	predictable with no fun

Explain how Naïve Bayes Classifier predicts the class for the movie review: "predictable with no fun".

6	Explain Named entity recognition as sequence labelling.	(3)
7	Show how the cosine computes which of the words "cherry" or "digital" is closer in meaning to information using raw counts from the following shortened table:	(3)

	pie	data	computer
cherry	442	8	2
digital	5	1683	1670
information	5	3982	3325

8   Consider a tiny query against a collection of 4 nano documents shown below: (3) Assume all words in the following query and documents are down cased and punctuation is removed. Compute tf-idf values and rank the documents.

     Query   sweet love

     Doc1 :   Sweet sweet nurse! Love?

     Doc 2:   Sweet Sorrow

     Doc 3:   How sweet is love?

     Doc 4:   Nurse!

9   Explain different evaluation metrics used for evaluating factoid answers.   (3)

10   Explain Statistical Machine Translation.   (3)

PART B
Answer any one full question from each module, each carries 14 marks.

Module I

11   a)   Explain the applications of NLP.   (7)

    b)   Explain what a language is and the building blocks of Language.   (7)

OR

12   a)   Explain why NLP is considered a challenging task.   (14)

Module II

13   a)   Consider a scenario where you have built a restaurant recommendation system.   (7) Mention a stage in NLP Pipeline that you would use to measure the goodness of the model. Justify the goodness of the model both in research and industrial perspective.

    b)   With the help of a diagram, explain NLP Pipeline for ranking tickets in a ticketing   (7) system by Uber.

OR

14 a) Explain two architectural variants of Word2Vec approach with the help of a diagram. (14)

Module III

15 a) Explain training in Naïve Bayes Classifier. (14)

OR

16 a) Explain Stochastic Gradient Descent Algorithm in logistic regression with the help of an example. (14)

Module IV

17 a) Explain different approaches to relation extraction (14)

OR

18 a) Given below are some list of scenarios: (14)

- Locating your favourite products on Amazon
- Searching for any email in Outlook.
- Imagine you are hungry and want to make a delicious authentic paella.

Hint: Our life would probably not be as easy without all those search boxes that allow us to find all sorts of documents and information, be it written, audio, photos, videos, or something else.

Select an appropriate NLP system based on the problems listed above. How do we evaluate the performance of such systems?

Module V

19 a) Explain IR-based Factoid Question Answering. (14)

OR

20 a) Explain the reasons why Machine translation is so hard. (14)

****