

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

School of Computing Science and Engineering**B.TECH CSE Artificial Intelligence and Machine Learning****Mid Term Examination - Nov 2023****Duration : 90 Minutes****Max Marks : 50****Sem V - E2UD501B - Natural Language Processing**General Instructions*Answer to the specific question asked**Draw neat, labelled diagrams wherever necessary**Approved data hand books are allowed subject to verification by the Invigilator*

- 1) Consider the following two expressions, which have the same value. Which one will typically be more relevant in NLP? Why? a. "Monty Python"[6:12] b. ["Monty", "Python"] [1] K3 (6)
- 2) Explain how NLP plays a role in the following areas. A) Image recognition B) Sound processing C) Human language understanding D) Data visualization K3 (9)
- 3) Given two binary word vectors w_1 and w_2 as follows: $w_1 = [1010101010]$ $w_2 = [0011111100]$. Compute the Dice and Jaccard similarity between them. K4 (8)
- 4) How do we convert text to feature using a count vectorizer? Write a python code for the same. K5 (15)
- 5) You are building a model distribution for an infinite stream of word tokens. You know that the source of this stream has a vocabulary of size 1000. Out of these 1000 words you know of 100 words to be stop words each of which has a probability of 0.0019. With only this knowledge what is the maximum possible entropy of the modelled distribution. (Use log base 10 for entropy calculation) [K6 (12)