ourdiu.com | Shared by Abir Mia -231

Daffodil International University

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Class Test - 1, Summer-2026

Course Code: CSE432 **Course Title: NLP**

Time: 45 minutes **Full Marks: 15**

Q1.	Consider the following three documents: D1: "data science is fun" D2: "data science is hard" D3: "football is fun" (a) Compute the Cosine Similarity between: • D1 and D2 • D1 and D3 (b) Which pair of documents is most similar according to Cosine Similarity?	4
Q2.	Given the following corpus of 4 documents: D1: "the cat sat on the mat" D2: "the cat sat on the hat" D3: "the dog sat on the log" D4: "the bird flew over the mat" (a) Calculate the TF-IDF score for **"cat", "mat", and "sat"** in document **D1.** (b) Which word gets the highest TF-IDF score in D1?	4

Q3.	A spam classifier is trained on the following dataset: D1: "win free money now" → Spam D2: "free lottery win prize" → Spam D3: "meeting at noon today" → Not Spam D4: "lunch meeting tomorrow noon" → Not Spam D5: "win the lottery today" → Spam Test Document: **"win free lunch"** (a) Calculate the Prior Probabilities.	5

Scanned with CamScanner

ourdiu.com | Shared by Abir Mia -231

	(b) Calculate the Likelihood of each word in the test document for both classes using Laplace Smoothing. (c) Which class does the classifier assign to the test document?	
Q4.	What is Maximum Likelihood Estimation (MLE) in the context of Bigram models? Write the MLE formula for computing a Bigram probability and explain each component.	2

ourdiu.com

Scanned with CamScanner