



**Daffodil International University**  
**Department of Computer Science and Engineering**  
**Faculty of Science & Information Technology**

Class Test - 1, Summer-2026

**Course Code: CSE432**

**Time: 45 minutes**

**Course Title: NLP**

**Full Marks: 15**

|     |                                                                                                                                                                                                                                                                                                                                                                          |   |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|
| Q1. | <p>Consider the following three documents:<br/> D1: "data science is fun"<br/> D2: "data science is hard"<br/> D3: "football is fun"</p> <p>(a) Compute the Cosine Similarity between:<br/> • D1 and D2<br/> • D1 and D3</p> <p>(b) Which pair of documents is most similar according to Cosine Similarity?</p>                                                          | 4 |
| Q2. | <p>Given the following corpus of 4 documents:<br/> D1: "the cat sat on the mat"<br/> D2: "the cat sat on the hat"<br/> D3: "the dog sat on the log"<br/> D4: "the bird flew over the mat"</p> <p>(a) Calculate the TF-IDF score for "cat", "mat", and "sat" in document D1.</p> <p>(b) Which word gets the highest TF-IDF score in D1?</p>                               | 4 |
| Q3. | <p>A spam classifier is trained on the following dataset:</p> <p>D1: "win free money now" → Spam<br/> D2: "free lottery win prize" → Spam<br/> D3: "meeting at noon today" → Not Spam<br/> D4: "lunch meeting tomorrow noon" → Not Spam<br/> D5: "win the lottery today" → Spam</p> <p>Test Document: "win free lunch"</p> <p>(a) Calculate the Prior Probabilities.</p> | 5 |

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