

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

Time: 50 min **Quiz-2**

Semester: Fall 2025

Course Code: CSE445

Course Title: Natural Language Processing

1. Consider the corpus:

D1: "data data science is fun";

D2: "science is fun fun";

D3: "data science is boring";

D4: "data data data is science";

D5: "fun is fun".

(a) Compute term frequencies for each document and inverse document frequencies for the corpus vocabulary.

(b) Compute the TF-IDF vector of each document.

(c) Identify the most informative term(s) in each document.

[7.5]

2. The film "Jawan" released in 2024. The following training set contains reviews with labels. Apply Laplace smoothing with $\alpha = 1$.

[7.5]

Review	Label
Jawan is great great	Good
Jawan is bad	Bad
Jawan is average	Neutral
great movie	Good
boring and bad	Bad
average is average	Neutral

A new review appears: "Jawan is not a great movie". Determine the predicted label. Show the full calculation steps.

-6 -5