ourdiu.com | Shared by Abdullah Tawsif (CSE-232)

Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science and Engineering
Midterm Examination, Fall - 2024
Course Code: CSE445, Course Title: Natural Language Processing
Level: 3 Term: 2 Batch: 61
Time: 1.5 Hours **Marks: 25**

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1.	a)	**Explain** the challenges of natural language processing.	[2]	CO1
	b)	**Explain** with example: Why NER is useful?	[3]	
2.		Consider the following statements: D1: "I love soccer a lot." D2: "Basketball is fun." D3: "Many love soccer and basketball." D4: "Ronaldo plays soccer."		CO2
	a)	**Calculate** TF, IDF and TF-IDF of for the word "Love" and "soccer" based on D1.	[2]	
	b)	**Find** the cosine similarity between D1 and D2, D1 and D3, and D1 and D4.	[3]	
3.		Consider the following text dataset:		CO2

	Comments	Class
Training	The laptop battery life is terrible	Negative
	This phone has an excellent camera	Positive
	The customer service was not great	Negative
	Fast delivery and great packaging	Positive
	The software is slow and crashes often	Negative
Test	Customers are not happy.	?

	a)	**Apply** the following preprocessing: i) Convert the corpus into small letter ii) Tokenization iii) Remove stopwords (determinants, be verb, prepositions and conjunctions) iv) **Construct** the updated table.	[2]
	b)	**Determine** the class for the given test data.	[3]
4.		An autonomous vehicle company is testing its object detection system on a dataset of 1200 images, where 600 objects are correctly detected as pedestrians, 100 images are incorrectly classified as having pedestrians (but they are actually objects like trees or cars), 400 objects are correctly identified as non-pedestrians, and 100 images are incorrectly classified as non-pedestrians (but they actually contain pedestrians).	CO3
	a)	**Construct** a confusion matrix from the scenario.	[1]
	b)	**Determine** Precision, Recall, F1-scrore and Accuracy.	[4]
5.		Suppose that you have built two machine learning models. The first model gives 60% accuracy on training data. The second model gives 90% accuracy on training data but shows 60% accuracy on validation data.	CO3
	a)	**Analyze** the performance of the models.	[2]
	b)	**Recommend** solutions to these models with explanation if you agree that the performances of these two models are not satisfactory.	[3]