BRAC UNIVERSITY
Department of Computer Science and Engineering

Examination: Midterm | Semester: Summer 2023
Duration: 75 minutes | Full Marks: 30

CSE 440: Natural Language Processing II

Figures in the right margin indicate marks.

Name:	ID:	Section:

Answer all 3

1. [CO1]

A. **Explain** with examples: [4]
   a. Precision
   b. Micro-average F1
   c. Macro-average recall
   d. Statistical significance

B. You have two classifiers ($y = \sigma(\mathbf{w}^T x + b)$), classifier A with weight $w_A = [1\ 0\ 2]^T$ and bias $b_A = 1$; and classifier B with weight $w_B = [1\ 2\ 0]^T$ with bias $b_B = 0$. For one example $X = [1\ 1\ 0]^T$, predict y_A and y_B. Which classifier incurs lower cross entropy loss if X's original label is 1? **Calculate. Show your work.** [6]

2. [CO1]

A. Let's say we are working with Shakespeare's plays, and we have three plays in our hand: Anthony and Cleopetra, Julius Caesar, and Hamlet, with four key characters: Anthony, Brutus, Caesar and Calpurnia. These characters appear in the plays as many times given in table 1. Consider this as your bag-of-words. Now **build** a term-term co-occurrence matrix using this binary bag-of-words. **Explain** what each value in the term-term co-occurrence matrix means. [6]

B. We have a document that has the word 'shallow' in it 25 times and the document is 100 words long. If we have a total of 10,000 documents and 1,000 of them have the word 'shallow' in it, calculate the TF-IDF of 'shallow' in that document. [4]

3. [CO2]

A. Gender bias in word embeddings means some words are closer to male words and other words are closer to female words. For example, a poorly trained word vector may associate 'doctor' to 'man' and 'nurse' to 'woman'. If I tell you your embedding has this issue, how will you **test** it? How can you **solve** it? **Explain.** [4]

B. Explain the training and testing process of the Naive Bayes algorithm for sentiment classification. That is, you are given N sentences where some are positive sentiment sentences and some are negative, and you have to train a Naive Bayes model that will be able to classify a sentence. Then, you got a new sentence which you do not know whether it is positive and negative, and you need to figure out. How are you going to do it? [6]

Table 1

	Anthony	Brutus	Caesar	Calpurnia
Anthony and Cleopetra	16	13	7	0
Julius Caesar	12	9	21	4
Hamlet	0	1	1	0