Ahsanullah University of Science & Technology
Department of Computer Science & Engineering
Title: Soft Computing (CSE 4237)

Continuous Assessment

Instructions:

1. Put a cover page at the top and write student name, student ID, course name, course ID, submission date, and signature.
2. Answer script should be handwritten on A4 white paper.
3. Write down the page number at the bottom of every page of the answer script.
4. Put your signature on top of every page of the answer script.
5. Before uploading rename the PDF file as
    a. ***StudentID***
        eg. 160204001.pdf

Question:

Consider the following 4 documents:-

1. *A female tennis player in a black top playing tenis*
2. *A woman standing on ## a tennis court holding a racket*
3. *A female tenis $$ player preparing to serve the b@ll*
4. *@ woman is holding a tennis rakket on a court*

Now answer the following questions:-

1. Construct a vocabulary for the documents.
2. Represent these sentences using one-hot representation and construct document vectors for each document.
3. Construct TF*IDF representation of these documents.

Note: No need to filter stopwords or any sort of preprocessing to reduce the vocabulary size.