**COMSATS University Islamabad, Lahore Campus**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Course Title: | Introduction to Data Science | | | | Course Code: | CSC461 | Credit Hours: | 3(3,0) |
| Resource Person: | Dr. Muhammad Sharjeel | | | | Programme Name: | BSSE | | |
| Semester: | 5th | Batch: | FA21 | Section: | C | Max Marks: | 10 | |

**HAMNA ASHRAF SP20-BSE-047**

**Assignment 4**

**Due Date: 11-12-2023**

Submission: Upload the assignment solution (PDF file and Python code, preferably iPython notebook) to your GitHub account (private repository).

Important instructions: Please write the following information at the start of your ipython file.

*# Date*

*# CSC461 – Assignment4 – NLP # Your Full Name*

*# You Complete Registration Number # A brief description of the task*

*Important Instruction:*

*Solve the following questions manually as well as implement the solution using Python. Submit both.*

Q1. Compute BoW, TF, IDF, and then TF.IDF values for each term in the following three sentences.

S1: “data science is one of the most important courses in computer science” S2: “this is one of the best data science courses”

S3: “the data scientists perform data analysis”

Q2. Compute the similarity between S1, S2, and S3 using cosine, manhattan, and euclidean distances.









