

National University of Computer and Emerging Sciences, Lahore Campus



Course:	Natural Language Processing	Course Code:	CS4063
Program:	BS(Computer Science)	Semester:	Fall 2021
Due Date	30-Dec-2021	Total Marks:	10
Type:	Assignment 3		
Sections	CS-9A		

Important Instructions:

1. Submit your source files in a zipped file named as your roll number, i.e., 20L-1111.zip.
2. You are not allowed to copy solutions from other students. We will check your code for plagiarism using plagiarism checkers. If any sort of cheating is found, negative marks will be given to all students involved.
3. Late submission of your solution is not allowed.

Emotion Detection

You are given a dataset about emotion detection. Your tasks are the following:

1. Perform the necessary preprocessing on the dataset that we have in the class.
2. Now, encode the data using TF-IDF vectors and word embeddings. You can use a library for this task. However, if you encode TF-IDF vectors from scratch, then you will get bonus marks.
3. Now divide the dataset into training set and test set.
4. Now train logistic regression, support vector machine and artificial neural network on both encodings. You will have 4 models in total.
5. Now test the models using the test set and find accuracy, F1-score, precision, and recall.

Thank You and Best of Luck.