COMP 5525 NLP – Assignment
Assoc. Prof. Dr. Mete EMİNAĞAOĞLU

Submission Deadline: 21st of December, Sunday at 21:00
(If you don't deliver your assignment on time, you will get 0)

You must deliver your report, and all of the source code of your program in a zipped file by uploading to this course's Moodle web portal's assignment section.

Assignment Materials, Methods & Rules:

You will use the datasets named as "train.csv.zip" and "test.csv.zip" for sentiment analysis by using the methods and constraints explained in the following sections.

- The datasets are Turkish sentences and their sentiment labels (Positive, Negative, Notr).
- You must use train.csv to train your model and then test your model with test.csv.

- **You must use two different algorithms and related models / methods and compare the results.** For instance, one of them could be tf-idf and k-NN and the other can be Multinomial Naïve Bayes. Or, one of them can be vector embeddings with Word2Vec, FastText, Glove, etc. with an Artificial Neural Network, and the other one can be binarized Multinomial Naïve Bayes. Or, one of them can be vector embeddings with Word2Vec, FastText, Glove, etc. with k-NN, and the other can be tf-idf and k-NN, etc. And so on…
- **You cannot use any LLM in this assignment, if you use any LLM, you get 0**.
- The **comparative performance scores and the algorithms, methods, parameters, etc.** must be included and delivered as a report. **If you don't include these in your report, you will be punished with a -40 from this assignment's grade. ATTENTION! Even your code generates these scores and displays on the screen as output, it is not accepted and you will still be punished with -40. YOU MUST provide the report with the table.**

(A sample for the report format & table)

Train results:

1st Algorithm name / model / methods/ parameters: … … …	2nd Algorithm name / model / methods/ parameters: … … …
Micro Average	Micro Average
Recall: 0.xxx	Recall: 0.xxx
Precision: 0.yyy	Precision: 0.yyy
F-Score: 0.zzz	F-Score: 0.zzz
Macro Average	Macro Average
Recall: 0.xxx	Recall: 0.xxx
Precision: 0.yyy	Precision: 0.yyy
F-Score: 0.zzz	F-Score: 0.zzz
Class: Positive	Class: Positive
Recall: 0.xxx	Recall: 0.xxx
Precision: 0.yyy	Precision: 0.yyy
F-Score: 0.zzz	F-Score: 0.zzz
Class: Negative	Class: Negative
Recall: 0.xxx	Recall: 0.xxx
Precision: 0.yyy	Precision: 0.yyy

F-Score: 0.zzz	F-Score: 0.zzz
Class: Notr	Class: Notr
Recall: 0.xxx	Recall: 0.xxx
Precision: 0.yyy	Precision: 0.yyy
F-Score: 0.zzz	F-Score: 0.zzz

Test results:

1st Algorithm name / model / methods/ parameters: … … …	2nd Algorithm name / model / methods/ parameters: … … …
Micro Average	Micro Average
Recall: 0.xxx	Recall: 0.xxx
Precision: 0.yyy	Precision: 0.yyy
F-Score: 0.zzz	F-Score: 0.zzz
Macro Average	Macro Average
Recall: 0.xxx	Recall: 0.xxx
Precision: 0.yyy	Precision: 0.yyy
F-Score: 0.zzz	F-Score: 0.zzz
Class: Positive	Class: Positive
Recall: 0.xxx	Recall: 0.xxx
Precision: 0.yyy	Precision: 0.yyy
F-Score: 0.zzz	F-Score: 0.zzz
Class: Negative	Class: Negative
Recall: 0.xxx	Recall: 0.xxx
Precision: 0.yyy	Precision: 0.yyy
F-Score: 0.zzz	F-Score: 0.zzz
Class: Notr	Class: Notr
Recall: 0.xxx	Recall: 0.xxx
Precision: 0.yyy	Precision: 0.yyy
F-Score: 0.zzz	F-Score: 0.zzz

- Students can use any of the programming languages and any built-in libraries, functions, etc. within that language given below (but you cannot use anything else besides the ones given below).
  C, C++, C#, .Net, Java, Python.

Mandatory Deliverables & Report:

1-All of your source code including the libraries, etc.
2-Necessary short notes & explanations in the source code as comment lines.
3-Comparative performance analysis table.