



## Exam 1

- **Show ALL Work, Neatly and in Order.**
- **No credit for Answers Without Work.**
- Open Books, Open Notes.

### 1 Introduction

This exam consists in using NLP techniques to classify a given text to its class. The dataset includes some textual information as a Text and encoded labels as Target. The goal of this exam is to use your NLP knowledge and data preprocessing that you learnt over time to perform text classification.

The format will be as following:

- Write a training script that could classify each text.
- Train and Test data files are given.
- F1-score as our metric for evaluation.

### 2 Dataset

- Download the exam Train.csv and Test.csv from BB Test section Exam 1.

### 3 Rules of Exam

Please read these rules **carefully** and if you have any questions please let me know directly.

- You can use Sklearn and any other packages for training except **gensim, transformers architecture and any neural networked based models. You are just allowed to use Naive bayes or Logistic regression models.**
- You can use Pandas, numpy, matplotlib, seaborn, scipy, matplotlib in this competition.
- You can only use the data you are given. Using additional data from any other sources is not allowed.
- You can do any kind of pre-processing with the **training data**, which you should split into at least training and testing. You may use whichever library you want for this purpose.

- You are not allowed to share your results with others. If we find out you will get **zero** grade for the Exam.
- You are not allowed to copy code or ideas from any students in the class. If we find out you will get **zero** grade for the Exam.
- You are allowed to search in the internet and find out ideas. You can use any external GitHub but you need to **cite** it. If we found any violation of this rule you get a reduce grade.

## 4 Clarifications on preprocessing

- You are allowed to do any pre processing on training and test set.

## 5 Deliverables

1. A training code and all of its subroutine if you have one.
2. A training code should print the f1 score on the test dataset.
3. A csv file which filled all the labels on the Target column of the test data.

## 6 Grading

1. Your F1 score and your feature engineering is going to be evaluated with other classmates.