Exam 2

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

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

1 Introduction

In this exam, you need to use NLP techniques that you learnt over the course to classify a text which has multiple labels. The dataset contains 6 different labels(Computer Science, Physics, Mathematics, Statistics, Quantitative Biology, Quantitative Finance) to classify the research papers based on their Title. The goal of this exam is to use your NLP knowledge and data preprocessing that you learnt over time to classify these review. You will be given a training set and then asked to submit a submission file, so that we can test it on our private held-out set. We will use the Sum of accuracy and Quadratic Cohen Kappa Score as a metric.

2 Dataset and Submission file

- Download the exam dataset from BB Test section Exam 2.

- Please check the sample submission file called ***Test_submission_nedid.csv***. You need to fill the label column with predicated values from your model. Please make sure you use the same name as your test file and the name of the file is case sensitive.

3 Exam Models

- **USE TRANSFORMER MODEL AS YOUR BASE MODEL. OTHER MODELS ARE NOT GRADED.**

- **ANY BASE TRANSFORMER ARCHITECTURE IS ALLOWED TO INSTANTIATE.**

- **THREE CLASSIFICATION HEADS NEED TO BE CODED AND ATTACHED TO THE BASE MODEL TO SOLVE THE EXAM**

- **NO TRAINER CODE FROM HUGGING FACE IS ALLOWED. CHECK THE LINK HERE**

  1. **FULLY CONNECTED LAYER(LINEAR)**
  2. **LSTM**
  3. **CONVOLUTION**

4 Rules

- You can only use the data you are given. Using additional data from any other sources is not allowed.

- 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. Also, any external GitHub can be used. Make sure that you **cite** all the sources, failure to that result in a reduced grade. If we found any violation of this rule you get a violation.

5 Clarifications on preprocessing

- You can use any nlp packages as needed for preprocessing.

- 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 allowed to do any pre processing on training set.

- **Note**: Makes sure you preserve the order of the submission file because the automatic grader code reads the instances based on the same order which is in the test file.

6 Deliverables

1. A single submission file `Test_submission_netid.csv` needs to be posted to BB.

2. A training code and all of its subroutine if you have one.

3. You need to submit item 1 and item 2 to Blackboard.