ELSEVIER

Scopus®

Take-Home Midterm Exam (2023/2)
Paper Classification (10% + 5%)

2110446 Data Science and Data Engineering

Released Date: 26th Feb 2024
Kaggle Submission Deadline: 11st Mar 2024
Package Submission Deadline: 18th Mar 2024

Objective

Please join Kaggle competition ASAP:
https://www.kaggle.com/t/ab4bf180150b4f1a8232af6e212a3667

- Be able to perform a data preprocessing (pandas) to prepare & clean data
- Be able to perform a machine learning model (model development & model evaluation)
  - Traditional classification
  - Deep learning model from HuggingFace
- Be able to perform a self-study on an advanced topic "multi-label text classification"
  - You can find example codes in the course GitHub and slide here.

6-1. Text Classification (TF-IDF): Open in Colab

6-2. Text Classification (BERT): Open in Colab

This is an individual take-home midterm exam.

- **<u>Text classification</u>**
- 18 classes of subject areas from Scopus

- **<u>Scoring</u>**
- 10% for the Kaggle result
- 5% for the submission package

Take-home midterm exam (10%): Paper classification

Invitation link to join competition:
https://www.kaggle.com/t/ab4bf180150b4f1a8232af6e212a3667

2110446 Data Science and Data Engineering Tools

Take-home midterm exam with Scopus Dataset

Host | Overview | Data | Discussion | Leaderboard | Rules

Off to a great start!
You've completed 7 of 9 tasks to launch your competition.

View Launch Checklist

Overview

This project is the individual take-home midterm exam, which is a part of the 2110446 Data Science and Data Engineering Tools (2023/2).

Timeline

Set Competition Deadline

Competition Host
ARPANANT SAENG-XUTO

Prizes & Awards
Kudos
Does not award Points or Medals

Participation
0 Competitors
0 Teams
0 Entries

Data set

- 18 classes of subject areas (CPX)
  - **CE** - Civil Engineering
  - **ENV**- Environmental Engineering
  - **BME** - Biomedical Engineering
  - **PE** - Petroleum Engineering
  - **METAL**- Metallurgical Engineering
  - **ME** - Mechanical Engineering
  - **EE** - Electrical Engineering
  - **CPE** - Computer Engineering
  - **OPTIC** - Optical Engineering
  - **NANO** - Nano Engineering
  - **CHE** - Chemical Engineering
  - **MATENG** - Materials Engineering
  - **AGRI** - Agricultural Engineering
  - **EDU** - Education
  - **IE** - Industrial Engineering
  - **SAFETY** - Safety Engineering
  - **MATH** - Mathematics and Statistics
  - **MATSCI** - Material Science

- Train 454 samples
- Test 151 samples

- Please note that the testing data is already cleaned, while the training data set is not preprocessed yet. You need to clean the training data (a manual process may be helpful).

Kaggle Scoring Criteria (10%)

There will be also a winner for technical aspect regardless of the rank (full 10 points)

- 5% from your private score
  - Macro F1 >= 0.65: 5 points
  - Macro F1 >= 0.60: 4 points
  - Macro F1 >= 0.55: 3 points
  - Macro F1 >= 0.50: 2 points
  - Macro F1 >= 0.45: 1 points
  - Otherwise: 0 point
- 5% from your private rank
  - Percentile [1 - 10] : 5.0 points
  - Percentile [11 - 20] : 4.5 points
  - Percentile [21 - 40] : 4.0 points
  - Percentile [41 - 60] : 3.5 points
  - Percentile [61 - 80] : 3.0 points
  - Percentile [81 - 100] : 2.5 points

Important Rules!!!

1) The result must be able to replicated by the code you submitted. Your code must be able to run your TA's. Otherwise, your project can't be graded.

2) Copying is prohibited 🚫

Submission+Report (5%)

- Project1
  - Part1: (1) source code, (2) uploaded csv (result), (3) your prepared data, (4) model weight
    - Please note that your result must be able to reproduce and it must be similar (close) to the result on Kaggle.
  - Part2: Report (word & pdf)
    - Chapter 1: Introduction
    - Chapter 2: Data preparation (How to prepare data along with data statistics)
    - Chapter 3: Model (explanation of your model)
    - Chapter 4: Results (must contain captured screen on Kaggle)
    - Chapter 5: Discussion, e.g., error analysis and how to improve it
    - Chapter 6: Conclusion
- Submission:
  - MCV (Take-home midterm exam: report submission)
  - Deadline 18th Mar 2024