CS 488/508: Introduction to Data Mining

Data Mining Challenge

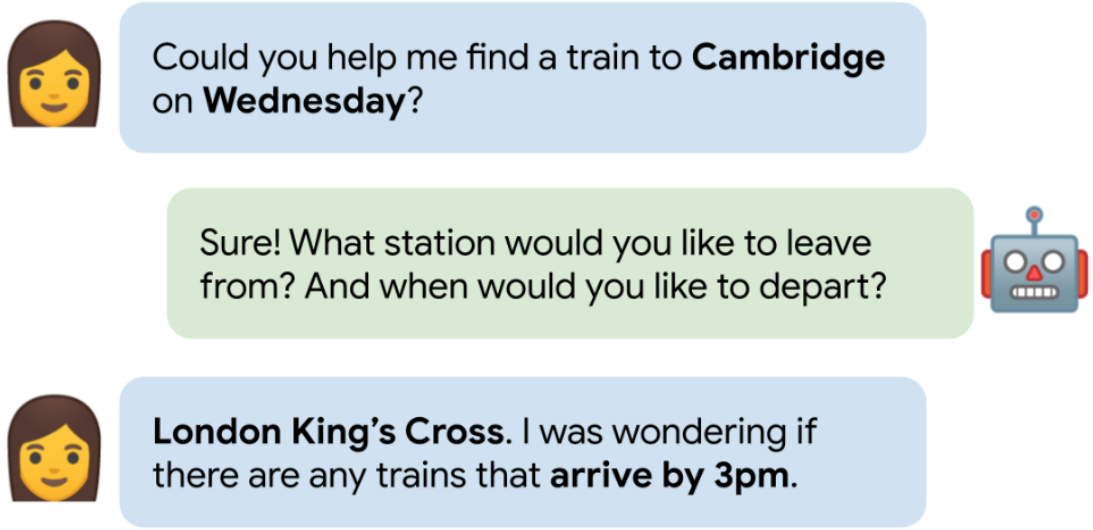
Tuan Le

College of Arts and Sciences | Computer Science

BE BOLD. Shape the Future.

Task-oriented dialog systems

- Task-oriented dialog systems have become ubiquitous, providing a means for billions of people to interact with computers using natural language.



https://blog.research.google/2022/04/simple-and-effective-zero-shot-task.html

Task-oriented dialog systems

- Tools for developing such systems start by guiding developers to collect training data for **intent classification**:

the task of identifying which of a fixed set of actions the user wishes to take based on their query

Challenge

- This is an optional competition. You don’t need to join the competition if you don’t want to.
- In this challenge, we focus on the task of **intent classification.** We will train a machine learning model to predict the intent of a query based on the insights that are gained from the training dataset.

dialog	intent
can you tell me how to say 'i do not speak much spanish', in spanish	translate
what is the equivalent of, 'life is good' in french	translate
can you check the status of my credit card application	application_status
how can i track the progress of my credit card application	application_status

Duration

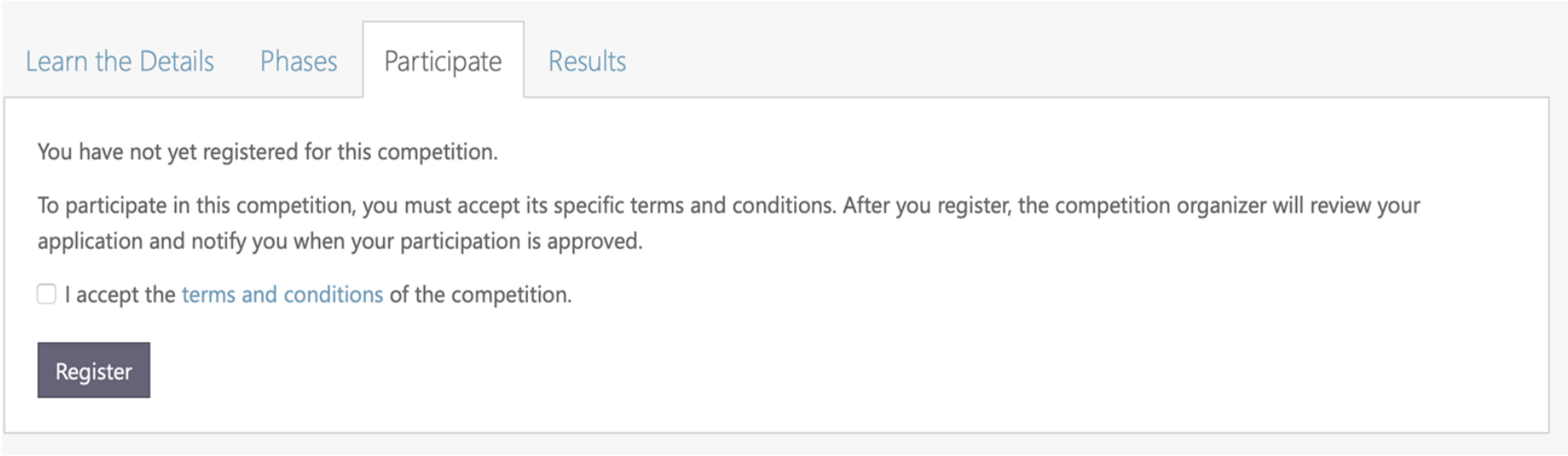
- Start: Oct 5, 9am, MT
- End: Nov 4, 5pm, MT

How to Join

- You can work individually or in a team (the teams can be the same as the project teams in class).
- Create a CodaLab account for your team at https://codalab.lisn.upsaclay.fr/ .
- **You must submit your team information here**: https://forms.gle/jLcQFAQ31Nqrwe4j9.
- Login and request to join the competition (see next slide) https://tinyurl.com/CS488508-DMChallenge2023

Participate

- Click on competition link: https://tinyurl.com/CS488508-DMChallenge2023
- Click tab “Participate”, you will see an interface showing similar to what follows.



Accept the terms and the conditions

Click the “Register” button

Data

- Data can be downloaded here: https://tinyurl.com/DMChallenge2023-data.
- Download data.zip, unzip it, you will see three files:
  - Training set: train.csv
  - Test set: test.csv
  - Example submission answer: answer.zip

How to Submit

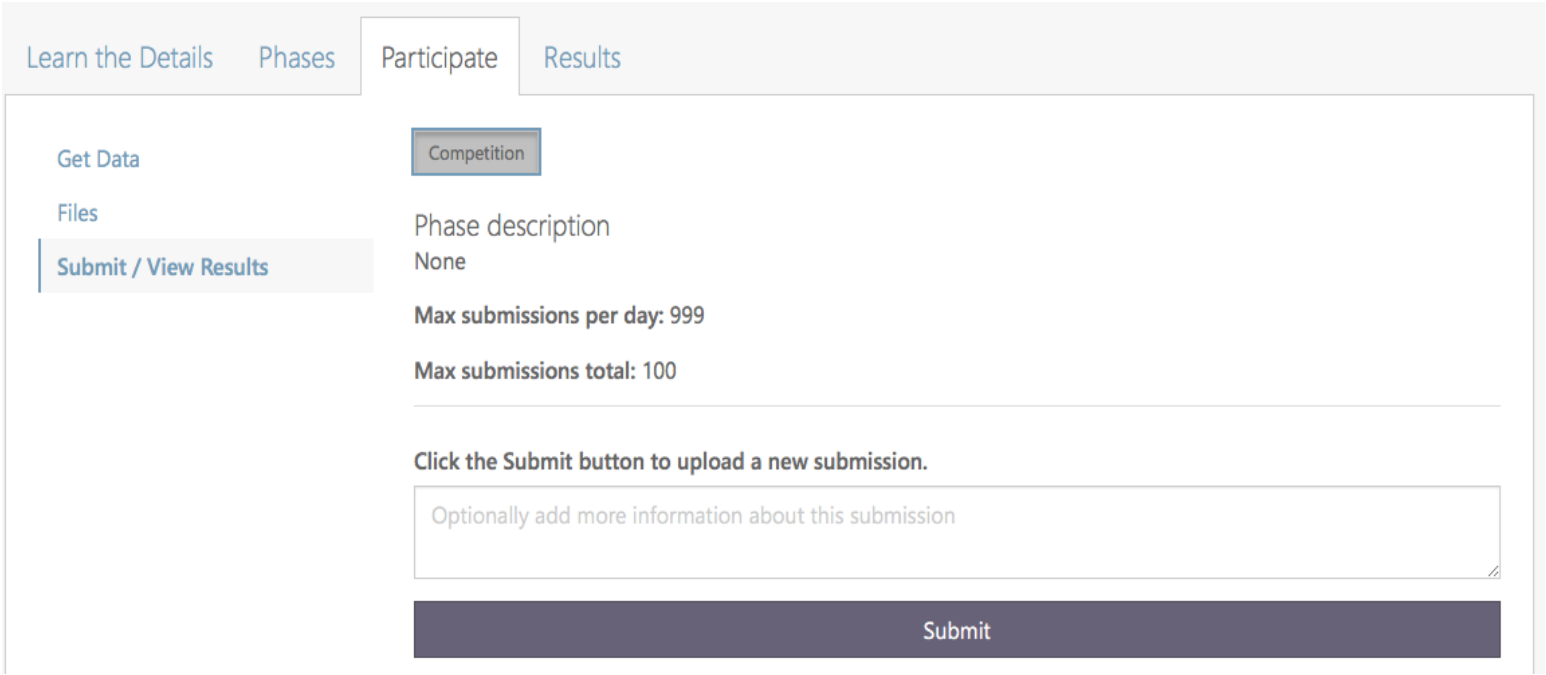
- Put your predicted labels for test.csv in answer.txt
  - One predicted label per line (see answer.zip for an example)

- Zip answer.txt (please zip answer.txt directly, don't put it in a folder), and submit the zip file.

- IMPORTANT:
  - Your answer file MUST still be named answer.txt.
  - You should only zip the answer.txt instead of a folder containing answer.txt
  - Your submitted file MUST be a zip file. It can be named as answer_team1_attempt2.zip or answer_team2_attempt2.zip, etc.

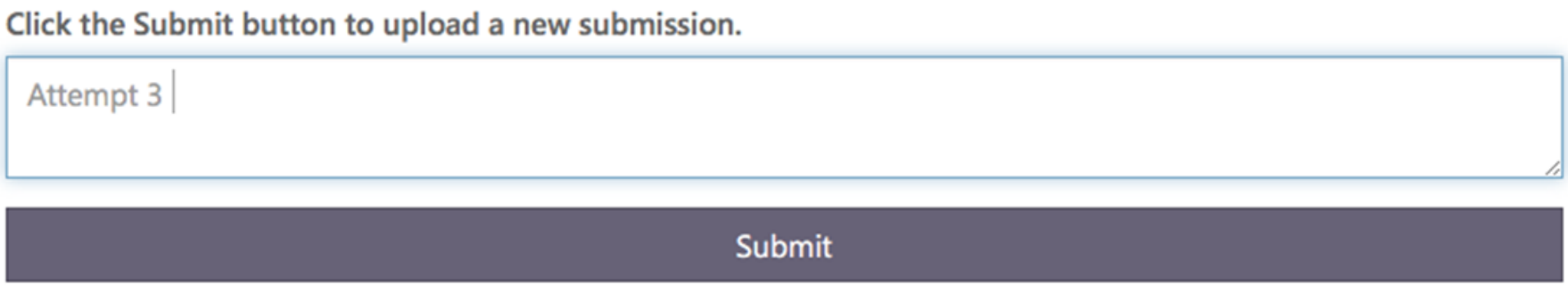
Submit your Answer (1)

- When you see the following interface, find and click “Submit/View Results” from the left panel.



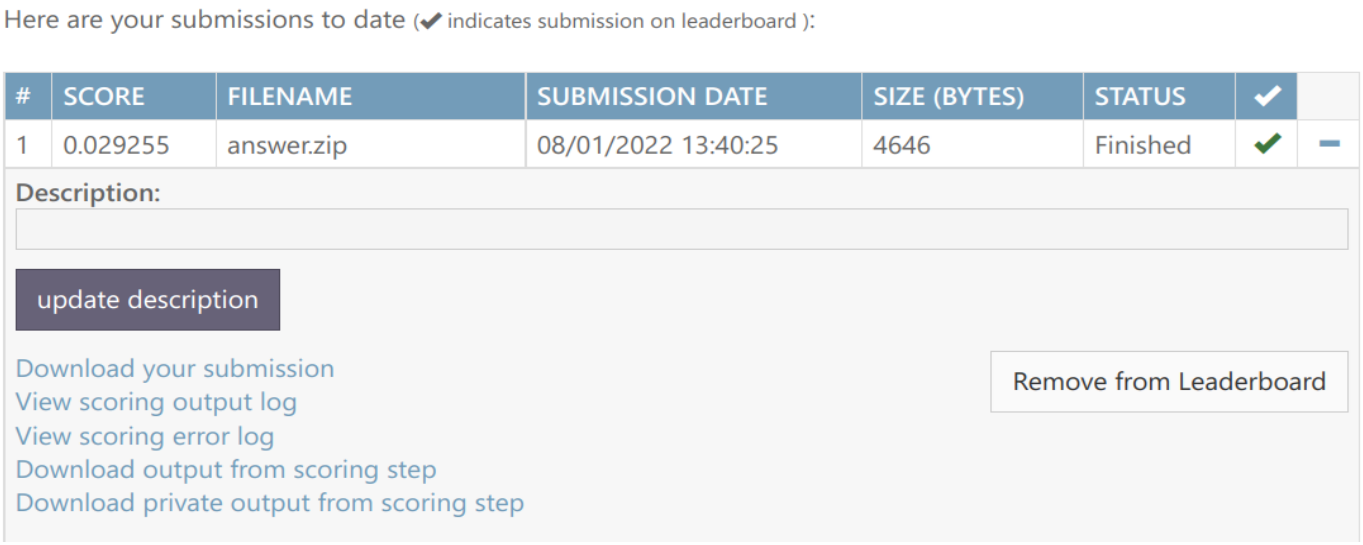
Submit your Answer (2)

- Put basic description about your submission in the box. Then, click the “Submit” button.
- After you click the Submit button, the system will prompt you to upload your answer file. Choose your answer_teamx_attempty.zip.



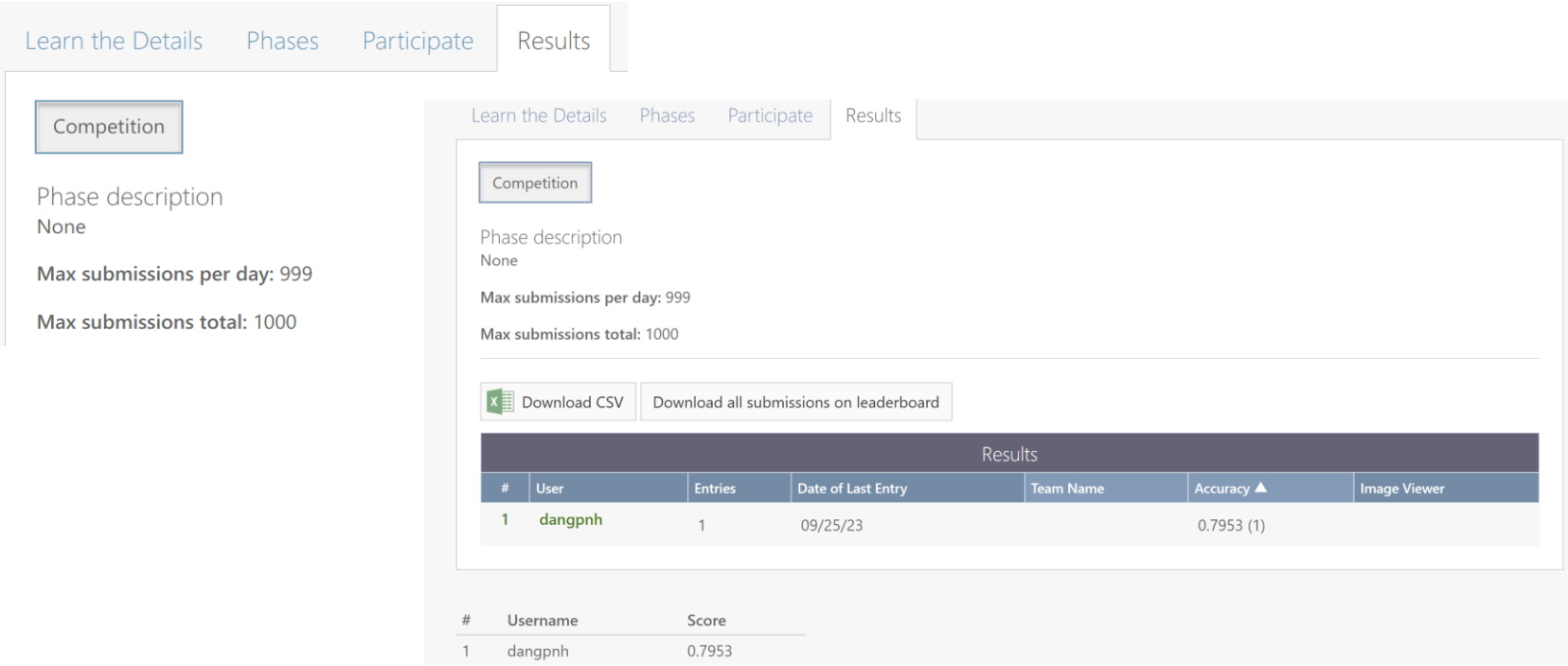
Submit your Answer (3)

- Now the dashboard should show your submitted files, submission dates, and status. Your recent submission status should be “Submitting”.
- Click the “Refresh status” button, your status will be changed to “Running” or “Finished”.



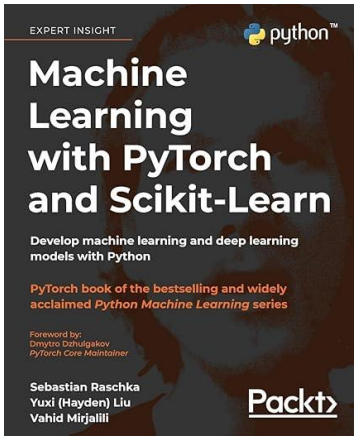
Results

- Click the “Results” tab, you should be able to see your best result and your rank. The maximum number of submissions is 1000. Do not submit more than 1000.



Prizes

- Prizes (sponsored by the CS Department)
  - 1st: one machine learning textbook per team
  - 2nd: one machine learning textbook per team



- Bonus points (sponsored by Dr. Le ☺):
  - Any team that achieves an accuracy >= 0.9 will get 100 points. This grade can be used to replace the lowest grade of **one** of your previous assignments. This bonus will apply to every member of the qualified team.

Enjoy and learn, Everyone!

Terms and Conditions

- Your team needs to work on the competition independently.
- The students are tasked to train a machine learning model to predict the values of the target variable in the test dataset based on the insights that are gained from the training dataset. Manually assigning the target values is not accepted.
- The winners are required to submit the source code/algorithms of their methods when the competition ends.

Brought to You by

- Computer Science Department

- Organizing Team:
  - Dr. Tuan Le
  - Phuong Nguyen
  - Dang Pham