Assignment

Text categorization: argument mining

1. Introduction

- Please read the document carefully and follow instructions. If you have questions ask in the telegram group.

- Use this template to complete your assignment report and upload it to Canvas: https://colab.research.google.com/drive/1ttPT6X4K0ovgbzmNjlcEiprkj1LaBuF2

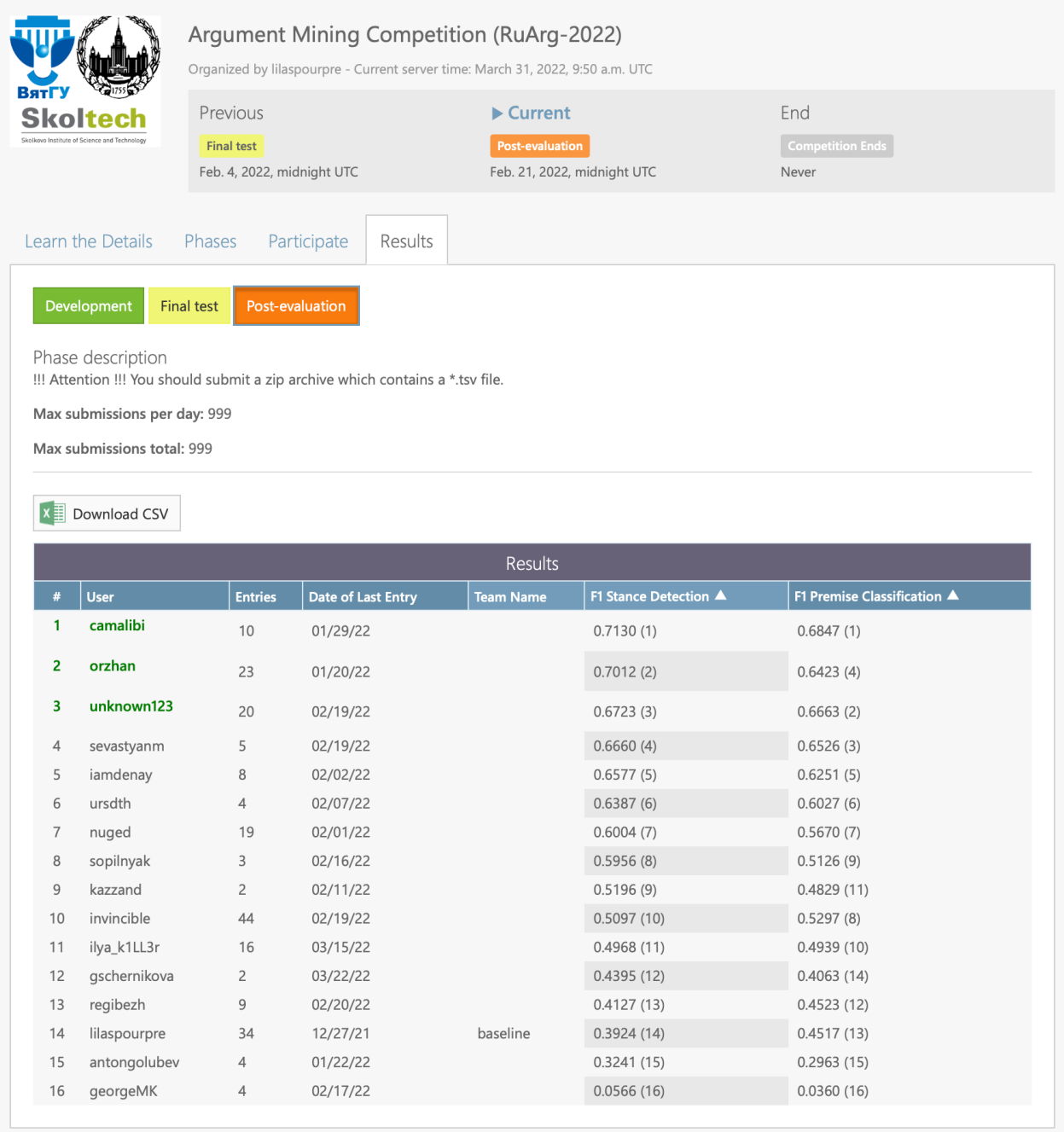
- Submit your results to Codalab (post evaluation phrase): https://codalab.lisn.upsaclay.fr/competitions/786
    - **Important: As the name of the team in Codalab indicate 'DL4NLP-23'.**
    - Important: In the Jupyter notebook indicate name of your Codalab user.

2. Task Description

In the context of this assignment, you will solve argument mining task as defined at the RuArg-22 competition at the Dialogue Conference. All details about the rules can be found at: https://www.dialog-21.ru/evaluation/2022/ruarg/, including the data, baselines, scripts, etc.

The shared task is a text categorization. To be more precise, you are asked to solve two text categorization tasks – stance detection and argument classification. The first one is about identification of the presence of a position of a person in a text with respect to the given topic. The second one is about presence or absence about argument in the text.

In this assignment, you are expected just to follow all the rules of the competition and prepare a fully-fledged submission to both tasks. A valid submission must appear in the 'Post-evaluation' tab on Codalab as show in the figure below:



Argument Mining Competition (RuArg-2022)

Organized by lilaspourpre - Current server time: March 31, 2022, 9:50 a.m. UTC

Previous	▶ Current	End
Final test	Post-evaluation	Competition Ends
Feb. 4, 2022, midnight UTC	Feb. 21, 2022, midnight UTC	Never

Learn the Details | Phases | Participate | Results

Development | Final test | Post-evaluation

Phase description

!!! Attention !!! You should submit a zip archive which contains a *.tsv file.

Max submissions per day: 999

Max submissions total: 999

Download CSV

Results

#	User	Entries	Date of Last Entry	Team Name	F1 Stance Detection ▲	F1 Premise Classification ▲
1	camalibi	10	01/29/22		0.7130 (1)	0.6847 (1)
2	orzhan	23	01/20/22		0.7012 (2)	0.6423 (4)
3	unknown123	20	02/19/22		0.6723 (3)	0.6663 (2)
4	sevastyanm	5	02/19/22		0.6660 (4)	0.6526 (3)
5	iamdenay	8	02/02/22		0.6577 (5)	0.6251 (5)
6	ursdth	4	02/07/22		0.6387 (6)	0.6027 (6)
7	nuged	19	02/01/22		0.6004 (7)	0.5670 (7)
8	sopilnyak	3	02/16/22		0.5956 (8)	0.5126 (9)
9	kazzand	2	02/11/22		0.5196 (9)	0.4829 (11)
10	invincible	44	02/19/22		0.5097 (10)	0.5297 (8)
11	ilya_k1LL3r	16	03/15/22		0.4968 (11)	0.4939 (10)
12	gschernikova	2	03/22/22		0.4395 (12)	0.4063 (14)
13	regibezh	9	02/20/22		0.4127 (13)	0.4523 (12)
14	lilaspourpre	34	12/27/21	baseline	0.3924 (14)	0.4517 (13)
15	antongolubev	4	01/22/22		0.3241 (15)	0.2963 (15)
16	georgeMK	4	02/17/22		0.0566 (16)	0.0360 (16)

You are allowed to do – until the deadline (as indicated by the ‘Date of Last Entry’ column) – as many submissions as possible. The last submission will be taken into account. Note that, valid submissions will be using ‘Team Name’ column.

In terms of the methods, you are allowed to use any methods or models. In your report you must describe which method you used for submission and how did you tuned its meta-parameters.

You can start with implementing a baseline using BERT-based classifiers as provided in the seminar. To get more ideas on how other participant of the shared task solved read please report of organizers and participants available at https://www.dialog-21.ru/evaluation/2022/ruarg/.

You are allowed to experiments with additional data for pre-training and multi-tasking of the model, e.g. other text classification datasets with similar semantics (sentiment, argument mining, etc.).

Other line of improvements may come from careful meta-parameters turning of the base model. You also could test various architectures not only relying on the ‘vanilla’ bert-based solution. Finally, do not forget to look at the data/results and perform error analysis to try to gain insights on which examples are problematic in terms of the classification and how these could be fixed.