

Deep Learning for Natural Language Processing

Course logistics

Examples of NLP Tasks

What is the topic of this article?

Is this replica toxic?

What does the user ask for from a smart speaker?

Are these two texts on the same topic?

What grammatical
errors are in the text?

Where are the names of organizations in this text?

Which of several texts looks more
natural?

What could be the text with such a heading?

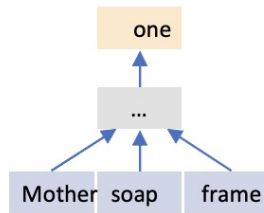
How can the bot answer the user if there
is no ready-made suitable answer?

What is the point of this article - in a couple of sentences?

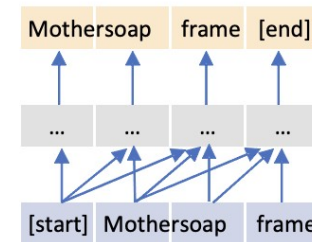
How would it be in Tatar?

Types of NLP tasks

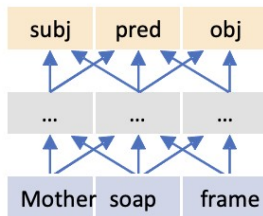
classification



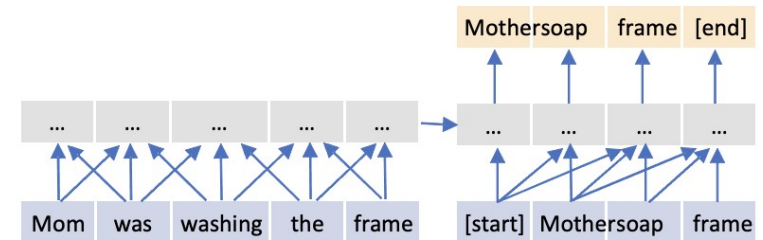
generation



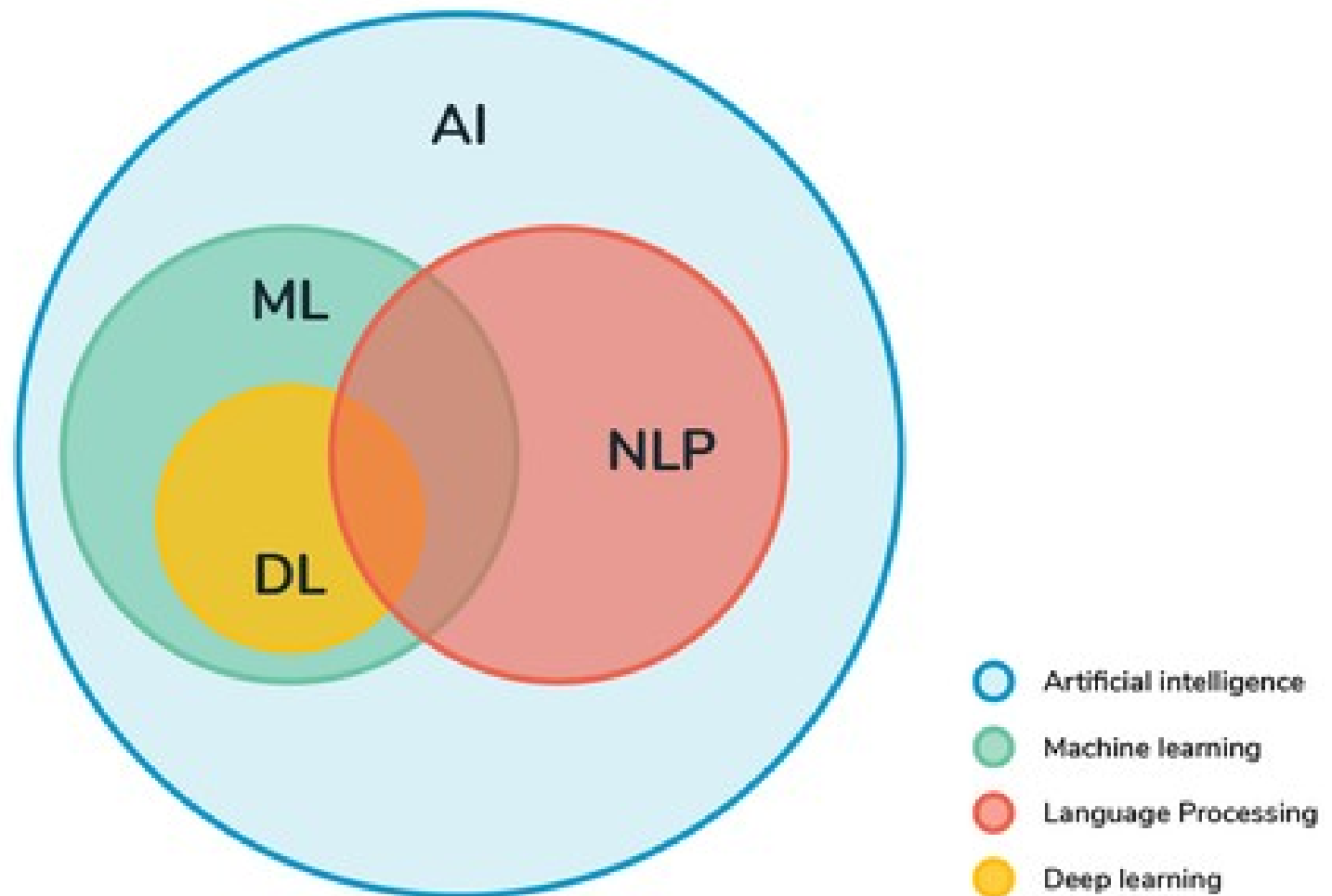
tagging



seq2seq



Deep Learning (DL) for Natural Language Processing (NLP)



This course is based and relies on multiple sources

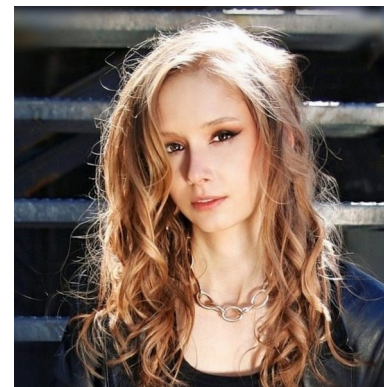
- Based on the materials from several courses:
 - “Transformers” at Sberbank Corporate University
 - “Neural Networks for Natural Language Processing” course by Nikolay Arefyev (Samsung Research Center & Moscow State University).
 - NLP course of Yandex School of Data Analysis.
 - Stanford course on “Natural Language Processing with Deep Learning” (cs224n).
 - Multiple other sources, indicating them where appropriate in the bottom of the slides.



Textbooks

- Yoav Goldberg (2017): Neural Network Methods in Natural Language Processing. Synthesis Lectures on Human Language Technologies.
<https://doi.org/10.2200/S00762ED1V01Y201703HLT037>
– Available at Skoltech library online
- Dan Jurafsky and James H. Martin (2022): Speech and Language Processing.
<https://web.stanford.edu/~jurafsky/slp3/>
- Lewis Tunstall, Leanadro von Werra, Thomas Wolf (2022): NLP with Transformers. O'Really.
<https://www.oreilly.com/library/view/natural-language-processing/9781098103231/>
- Kyunghyun Cho (2015): Natural Language Understanding with Distributed Representation. Lecture note for the course DS-GA 3001 on Natural Language Understanding with Distributed Representation at the Center for Data Science, New York University.
<https://arxiv.org/abs/1511.07916>
- Delip Rao and Brian McMahan (2019) Natural Language Processing with PyTorch. O'Really.
<https://www.oreilly.com/library/view/natural-language-processing/9781491978221/>
- Denis Rothman (2021). Transformers for Natural Language Processing. Packt.
<https://www.packtpub.com/product/transformers-for-natural-language-processing/9781800565791>

Team of instructors



Course logistics

Schedule of the course

Week	Date	Time	Session Type	Topic	Lecture	Seminar
1	28.03	16-19	Lecture+Seminar	Course logistics. Base neural architectures.	Alexander	Mikhail
	31.03	16-19	Lecture+Seminar	Text detoxification talk (TrueTechDay) + Embeddings	Alexander	Daniil
2	4.04	16-19	Lecture+Seminar	Motivation for Transformer. Attention. The original Transformer architecture.	Alexander	Maria
	4.04	EoD	Assignment 1 release	Transformer and transliteration seq2seq task	Irina	
	7.04	16-19	Lecture+Seminar	Motivation for Transformer. Attention. The original Transformer architecture (cont'd)	Alexander	Maria
	9.04	EoD	Assignment 1 deadline	Transformer and transliteration seq2seq task	Irina	
3	11.04	16-19	Lecture+Seminar	Transformer-based Encoders. Masked language models based on the Transformer architecture. BERT and related models.	Alexander	Daniil
	11.04	EoD	Assignment 1 deadline	Transformer and transliteration seq2seq task	Irina	
	11.04	EoD	Assignment 2 release	Text categorization and argument mining task	Irina	
	14.04	16-19	Lecture+Seminar	Transformer-based Encoders. Masked language models based on the Transformer architecture. BERT and related models (cont'd)	Alexander	Daniil
4	18.04	16-19	Lecture+Seminar	Classification and sequence tagging with Transformers. Using encoders to generate feature representation for NLU tasks.	Alexander	Mikhail
	21.04	16-19	Lecture+Seminar	Classification and sequence tagging with Transformers. Using encoders to generate feature representation for NLU tasks (cont'd).	Alexander	Mikhail
	23.04	EoD	Assignment 2 release	Text categorization and argument mining task	Irina	
5	25.04			Transformer-based Decoders. Generation of text based on the Transformer architecture. GPT and related decoders. Text generation methods. Prompt tuning.		
		16-19	Lecture+Seminar		Victoriia	Maria
	25.04	EoD	Assignment 3 release	Text detoxification and textual style transfer	Irina	
	28.04			Sequence to sequence tasks: machine translation, text detoxification, question answering, dialogue.		
		16-19	Lecture+Seminar		Alexander	Mikhail
6	12.05	16-19	Lecture+Seminar	Uncertainty estimation for Transformer-based NLP models	Artem S	Artem V
	14.05	EoD	Assignment 3 deadline	Text detoxification and textual style transfer	Irina	
7	16.05	16-19	Lecture+Seminar	Multilingual language models based on the Transformer architecture.	Alexander	Maria
	19.05	16-19	Lecture+Seminar	Network encoders with Transformers	Irina	Irina
8	23.05	16-19	Lecture+Seminar	Compression of transformer models. Efficient transformers.	Victoriia	Victoriia
	26.05	16-19	Lecture+Seminar	Multimodal transformer-based models for text and images	Anton	Anton
	28.05	EoD	Assignment deadline	End of course and all late submissions.	Irina	

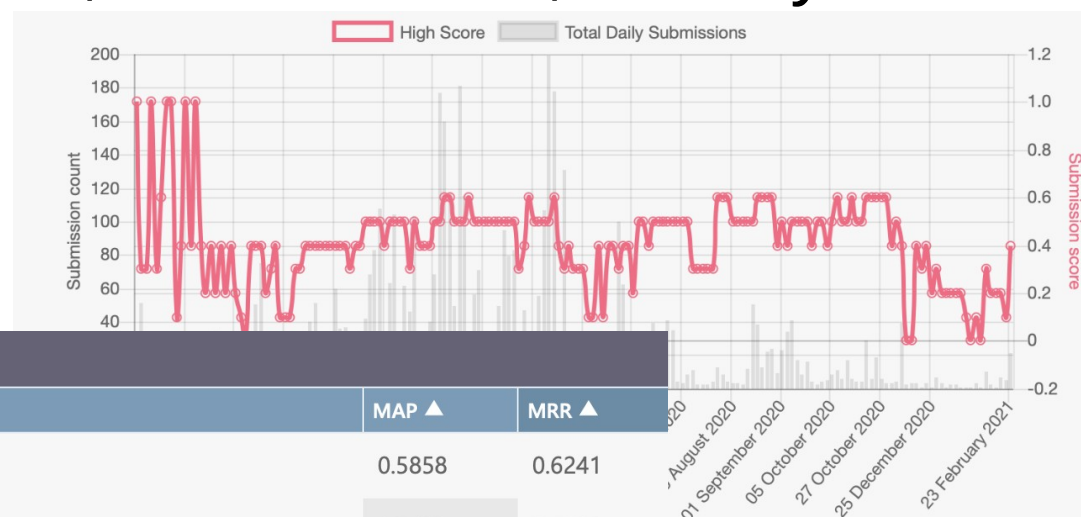
Assessment

- Three individual **assignments 33.33% each**
- You can also gain **extra points** (on top 100%) for
 - **Top results on the assignments:** top 1 or top 20% in leaderboard across all registered students
 - **Top 20% results on quiz questions** during lectures – worth 15% (1% per each of 15 sessions) across all registered students
 - No exam, but a final individual test may be arranged (**20% of extra points**)

Assignments

- A Kaggle-style competition for the best score
- Three tasks: transliteration, classification, and style transfer

CodaLab



RESULTS

#	User	Entries	Date of Last Entry	Team Name	MAP ▲	MRR ▲
1	RefalMachine2	334	08/28/20		0.5858	0.6241
2	vvyadrincev	25	03/17/20		0.5163	0.5506
3	mkunilovskaya	29	03/07/20	KuKuPI	0.4901	0.5301
4	aleksei_pronkin	9	11/23/20	Aleksei_Pronkin	0.4218	0.4521
5	vadim_bz	8	12/20/20		0.4210	0.4518
6	lilaspourpre	3	03/04/20	baseline (fasttext bugfixed)	0.4210	0.4518
7	iamdenay	7	03/12/20		0.4208	0.4512
8	Igor_Karpikov	2	11/23/20		0.0928	0.1019

Assignments – Grading policy

Technical report		Code		Results		Total
Methodology	Discussion of results	Readability	Reproducibility	Improved over the baseline	top-1 - 10 points top-20% - 5 points	100% + bonus
5	5	5	5	5	5 or 10	25 + 5 or 10

- Late submission:
 - There will be no deadline extension
 - 7 extra days are given to everyone for any assignment

Presentation of code and reports

- A single colab notebook following a pre-defined format

ASSIGNMENT-TEMPLATE.IPYNB

<> M

1. Information about the submission

1.1 Name and number of the assignment

1.2 Student name

1.3 Codalab user ID

1.4 Additional comments

2. Technical Report

2.1 Methodology

2.2 Discussion of results

3. Code

3.1 Requirements

3.2 Download the data

3.3 Preprocessing

Launcher

assignment-template.ipynb

Python 3

2. Technical Report

Use Section 2 to describe results of your experiments as you would do writing a paper about your results. DO NOT insert code in this part. Only insert plots and tables summarizing results as needed. Use formulas if needed to describe your methodology. The code is provided in Section 3.

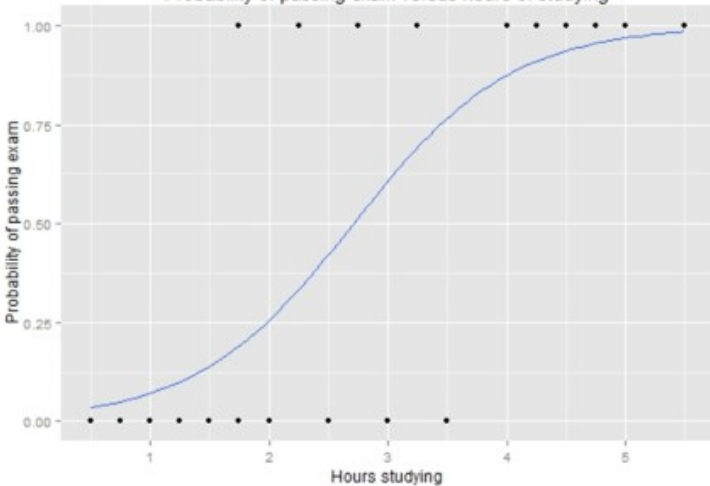
2.1 Methodology

Enter here a detailed description of the method used in your submission(s) to Codalab. The description should be at least 2-4 paragraphs featuring the following: type of the model, meta-parameters, how did you select meta-parameters, any further modifications of the out-of-the-box solutions, etc. The text is markdown and you can use math environment to write formulas:

$$\hat{y} = \beta_0 + \sum_{j=1}^p x_j \beta_j$$

Also you can insert images as needed:

Probability of passing exam versus hours of studying



Hours studying	Probability of passing exam
0	0.00
0	0.00
0	0.00
0	0.00
0	0.00
0	0.00
0	0.00
0	0.00
0	0.00
0	0.00
5	1.00
5	1.00
5	1.00
5	1.00
5	1.00

This part of the should contain description of all methods that you tried and, most importantly, that worked the best for you. Here you can include some tricks of your preprocessing, description of the models and motivation of their usage, the description of the training process details (train-test split, cross-validation, etc.). So, everything valuable that will help us to understand the scope of your work