

Intro to DL Y2022 Semester B

Project part 1 checklist

Checklist

1. Register on kaggle: <https://www.kaggle.com/>, save your username & password, you will use it during the entire semester. Your username should be your real name(s), this is the only way to ID you.
2. Register on Google colab <https://colab.research.google.com/notebooks/intro.ipynb> , save your username & password, you will use it during the entire semester.
3. Read sklearn introduction at <https://scikit-learn.org/stable/tutorial/index.html> . You can use any other sklearn tutorial you can find.
4. Read about nltk and tokenization here: <https://www.analyticsvidhya.com/blog/2019/07/how-get-started-nlp-6-unique-ways-perform-tokenization/> and here: <https://www.nltk.org/api/nltk.tokenize.html> .
5. Read how to represent text with the BOW model and sklearn: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html and https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html .
6. Download training, validation, and test data from moodle.
7. Write the assignment at colab.
8. Evaluate your solution using the training and the validation data.
9. Run your best code on the test data and store it in the required format.
10. Submit your results on kaggle competition site provided in the assignment description.