

Practice 8

Classical machine learning models for text classification

Commonly for text classification these models are using:

1. Multinomial Naïve Bayes (NB)
2. Logistic Regression (LR)
3. Support Vector Machines (SVM)
4. Stochastic Gradient Descent (SGD)
5. k-Nearest-Neighbors (kNN)
6. RandomForest (RF)
7. Gradient Boosting (GB)
8. XGBoost (the famous) (XGB)
9. Adaboost
10. Catboost
11. LigthGBM
12. Passive Aggressive Classifier

Task 1: choose 1 classical algorithm, make a report with code examples (5-10 pages) and presentation for 10-15 minutes.

Task 2: apply the chosen algorithm for true-fake news classification.

Literature:

1. <https://arxiv.org/pdf/2008.00364.pdf>
2. https://scikit-learn.org/stable/auto_examples/text/plot_document_classification_20newsgroups.html#sphx-glr-auto-examples-text-plot-document-classification-20newsgroups-py