Task 2: Sentiment analysis

Data

A subset of *Yelp reviews*: https://disk.yandex.ru/i/_OBHZ9GJ_nJOdg

Zhang, Xiang, Junbo Zhao, and Yann LeCun. "Character-level convolutional networks for text classification." *Advances in neural information processing systems* 28 (2015).

Methods

1. Lexicon-based using *SentiWords* https://disk.yandex.ru/d/iBRBoC6wQZkVyg

   Gatti, L., Guerini, M., & Turchi, M. (2015). SentiWords: Deriving a high precision and high coverage lexicon for sentiment analysis. *IEEE Transactions on Affective Computing*, *7*(4), 409-421.

2. *Stanza*'s sentiment model https://stanfordnlp.github.io/stanza/sentiment.html

Subtasks and points

1. Describe *Yelp* data. (15)
2. Process *SentiWord* data, describe the result. (20)
3. Develop a lexicon-based sentiment classifier using Stanza for lemmatization and POS-tagging. (Mind difference in labeling: sentences: 0 – negative, 1 – positive; words: continuous scores from the range [-1, 1]. Note that SentiWords and Stanza use different POS tag sets.) (35)
4. Evaluate the classifier on the Yelp data, report confusion matrix and F1 scores for each class (negative and positive). (10)
5. Apply Stanza's sentiment analysis model to the Yelp data, report confusion matrix and F1 scores for each class (negative and positive). (20)