Course project - IMDB movie rating

Task Description

IMDB dataset having 50K movie reviews for natural language processing or Text analytics. This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. It provides a set of 25,000 highly polar movie reviews for training and 25,000 for testing. So, predict the number of positive and negative reviews using either classification or deep learning algorithms.
For more dataset information, please go through the following link,
http://ai.stanford.edu/~amaas/data/sentiment/

The dataset can be found at http://ai.stanford.edu/~amaas/data/sentiment/aclImdb_v1.tar.gz

Requirements

Submit a notebook or pdf that containes every steps in data management, model building and outcome evaluation. Clearly describe the strategies used and explain the results.

The code should be uploaded to Github and the Github link should be provided in the report.

Reference

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. (2011). Learning Word Vectors for Sentiment Analysis. *The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011).*

Contact TA: Cai Mingxuan mcaiad@ust.hk