Character Level Models for Text Classification

TA: Kawshik Manikantan
Project ID: 8

This project involves a study regarding the classification of text using character-level models and also how they compare with other baseline models. You are required to perform an in-depth analysis of the baseline models as well as recent Character-level models. The analysis needs to point towards where exactly each model fails and measures that can be taken to improve their output.

The datasets(each corresponding to a unique text-classification task) to be used are

- Yelp Review Full
- dbpedia 14
- Amazon polarity

Requirements:

- Implement the following baseline models from scratch and improve their performance to the maximum extent possible:
    - BoW: Bag of Words
    - BoW: Bag of Words with TF-IDF
    - LSTM (LSTMCell of Keras allowed)
- You are also required to implement a Character Level Convolution Network (paper) from scratch and finetune CANINE: a character level Transformer (paper) on the three tasks
- Provide an intuitive and explanatory analysis of the results of the model. It is also required to make an effort to create explainable models.
- Develop an end-to-end system that takes as input a text and performs one of the tasks (with one of the models trained) on the custom input and provides the output.