Challenge

This task will test the following: Your understanding of coding styles and ability to quickly adapt code. Your ability to find your way around to solve new problems and to get familiar with the new tools. An opportunity to learn state-of-the-art embedding tools and their adaptation to task. The Google BERT, is the current state of the art in contextual word embeddings that offers both pre trained embeddings as well as fine-tuning to new scenarios. The link below is the repository for BERT:
https://github.com/google-research/bert
You are required to build a BERT based text classifier, by adding extra layers to the BERT embeddings second last layer [You will determine which kind of neural network layers to add].
Please use the following URL to download the text classification dataset
https://noisy-text.github.io/2017/emerging-rare-entities.html
In order to evaluate your model's performance, please the following official script to compute the F1 Scores on the test data:
wnuteval.py

What we are interested in?

- Your data processing decisions
- Your ability to understand, adapt and utilize the resources available for your task.
- Your coding ability.

Submission:

• Submit your code in a ZIP folder or a Github link .
• Please submit the prediction file (on test data) in CoNLL format (word, gold standard tag, predicted tag).
• Evaluation results in txt file on the test data using the aforementioned evaluation script.The two datasets are for the Named Entity Recognition (NER) task. Important to submit your work, you require a description of your system (An image and a paragraph), and the F1 score of your system. Please provide us your Github URL for the solution.
Good Luck!