

Task 1. Natural Language Processing. Named entity recognition IMPROVEMENTS

Of the major possible improvements to the NER model, it is worth highlighting such as:

- *Using better quality data:*
Expanding and improving the labelled dataset so that the model is trained on a larger and more diverse amount of data.
Pay attention to the variety of text styles and contexts in which mountains are mentioned. That is, create a broader version of the context for mentioning the name of a mountain.
- *Data augmentation techniques:*
Applying data augmentation techniques to increase the diversity of the training dataset. For example, using synonyms or changing sentence structure.
- *Better tuning of hyperparameters:*
Using such tools as Ray, optimize model hyperparameters such as learning rate, batch size, number of training epochs and other parameters to achieve better results.
- *Competition of multiple models:*
Using MLFlow to track and evaluate multiple NER models (e.g. SpaCy) or approaches; comparing their accuracy and stability within MLFlow client.
- *Additional Languages:*
A model capable of handling multiple languages can be applied in different contexts and for different language communities without the need to train a separate model for each language.

Conclusion:

Using LLMs, successfully gathered a dataset that includes most common mountain names on planet Earth. Then the NER model was trained for recognizing mountain names in texts and reached rather high performance on this dataset, this is great. However, given above recommendations for improving the model can be extremely useful for further improvement and extending its functionality.