Possible improvements

- Dataset

    Dataset was created using 2 already-made solutions: Few-NERD and NERetrive. Only instances that contain mountain entities were added to final version of dataset. So model worked only with examples that contain mountain names. It can has negative effect on model performance. As we can in demo.ipynb model constantly recognize mountain name that were used in figurative meaning.

    So the first improvement is to add examples without mountain entities or entities in figurative meaning.

- Model

    For this task fine-tuned BERT for named entity recognition was used. This was base BERT model. So second improvement is to check distil or large version. Also it is good idea to try other models, not BERT.

- Hyperparameters

    Hyperparameters were chosen based on original BERT paper. So third possible improvement is to perform hyperparameter search.