

Potential improvements

- The current training dataset is small (404 sentences) which can lead to underfitting. Increasing the number of training instances can improve model performance. In addition, It makes sense to do some web scrapping (from social medias, blogs, articles, news) in order to make the data more representative of real word cases.
- Another improvement is hyperparameter tuning which wasn't done in this project (learning_rate, batch size, etc.).
- In order to deal with the problem of names of mountains overlapping with other definitions (for instance, Olympus as a music band and name of the mountain) the BERT model can be tuned to capture contextual meanings of words. Another way would be putting additional labels to the words (like hiking, camping, climbing, etc.) that usually appear in sentences connected with mountains. Also It can be given a try to train the model with the data which contains sentences with words in both meanings (names of the mountains labeled as “-mount”, other meanings labeled as “O”)