| CURRENT RESULTS AND HOW TO MAKE THEM BETTER

1. Current results analysis

I used pre-trained BERT model. My dataset consists of ~730 sentences. 20 % of them were used for testing and evaluation of a model. Learning rate: 2e-4, number of training epochs: 5 and train and eval batch sizes were 16. I have achieved the following results (on test data) with this architecture:

```
[16] result

{'eval_loss': 0.03744079014846979,
 'precision': 0.9419642857142857,
 'recall': 0.9906103286384976,
 'f1_score': 0.965675057208238}
```

2. What can we do for better performance?

1. Add to the dataset more complex sentences with many mountain names. In the current dataset most sentences contain only one mountain. We also need to increase the size of the training sample to 1500-2000 sentences.

2. At the same time, the number of epochs should be increased. I was limited in GPU, so I trained the models on a small number of epochs. For the current dataset, it is enough to increase the number of epochs to 8-10, after this number, there is a tendency to overtraining

3. We can try to use larger pre-trained models. There is BERT-large (~3x bigger) and RoBERTa-large models (~3.5x bigger), but they take up much more space

4. If we have more computing power, it would be fine to increase bath sizes to 32+ (64, 128) size. It will give us more accurate estimates of the gradient during backpropagation.