

Report on the First Task

Kyrylo Krocha

December 8, 2024

My solution of the first task is based on fine tuning BERT model, and solution works properly, though maybe using better tuned models like RoBERTa could improve performance. But main problem is the training data - correctly labeled input is too small and I didn't have the time to expand it further. So I decided to ask ChatGPT to generate bio-formatted text fully on its own. The problem with this approach is that chatGPT sometimes labels countries as mountains. So putting better dataset for training the model should significantly improve the performance.