

How to improve the solution:

1. More data

The biggest issue with the solution is that dataframe size needs to be bigger for such a task. Therefore, to improve the performance more quality data should be gathered.

- a) Data should be domain-specific. If our aim is to classify mountain names in a scientific context, it would be a good idea to look for geological magazines and domain-specific articles to provide the model with more context.
- b) A larger amount of data should be collected. Previously, texts were collected by hand. Writing a web scraper to gather texts would be a good idea.
- c) Optimize the prompt. The prompt is too large. It could be reduced. The lesser the size of a prompt - the lesser the expenses for each API call. That means for each unit of resources (money) more data can be collected.