Report with Improvement Proposals for NER problem: Mountain Name Identification

Introduction

This project initially aimed to develop a Named Entity Recognition (NER) model focusing on the identification of mountain names within texts. Utilizing a pre-trained BERT model, the project involved fine-tuning with a custom dataset. The scope encompassed dataset creation, model selection, training, and demonstration via a Jupyter notebook.

Achieved results

The current model demonstrates commendable performance with standard benchmarks. However, there is a scope for improvement in complex contextual situations and varied sentence structures. The detailed metrics achieved are:

precision	recall	f1-score	support
0.98	0.98	0.98	692

Areas for Improvement

Enhanced BERT Architecture: Utilizing a more sophisticated BERT variant, such as BERT-Large or RoBERTa, could significantly enhance understanding of nuanced sentence structures and contexts, thereby improving entity recognition accuracy.

Dataset Development: Expanding the dataset is crucial. Proposed methods include integrating sentences with intricate structures and ambiguous contexts, challenging the model's comprehension capabilities. This approach aims to refine the model's performance in real-world scenarios where context plays a pivotal. However, creating such dataset is costly.