Mathematical Foundations of Data Science Course Project

Semantic Similarity

Libraries Used

I used sentence transformers, which provides an easy method to compute dense vector representations for **sentences**, **paragraphs**, and **images**. The models are based on transformer networks like BERT / RoBERTa / XLM-RoBERTa etc. and achieve state-of-the-art performance in various task. Text is embedding in vector space such that similar text is close and can efficiently be found using cosine similarity.

This framework provides an increasing number of **state of art pretrained models** for more than 100 languages, fine-tuned for various use-cases.

Further, this framework allows an easy fine-tuning of custom embeddings, to achieve maximal performance on your specific task.

Computed scores for each model combination and predicted the accurate one.