

Metrics:

Model	Dataset Used	Training Loss	Cosine Similarity
S-BERT	SNLI + MNLI	3.271592	0.9754

Challenges Faced:

1. The original BERT class does not support attention_mask
2. While populating the testing notebook to develop a POC for the web app, it was found out that inferencing couldn't be done on a system without Cuda support.

Mitigations Applied:

1. attention_mask was replaced it with segment_ids.
2. Inferencing was done in a GPU based system.

Proposed alternatives:

1. For a production level use case, the inferencing can be done by the HuggingFace inference API instead of a Ngrok makeshift solution currently implemented.