Models:

Length -> 128

CS -> Cosine Similarity (final result we got taking cosine similarity between embeddings adn then scaling to 1 to 4)

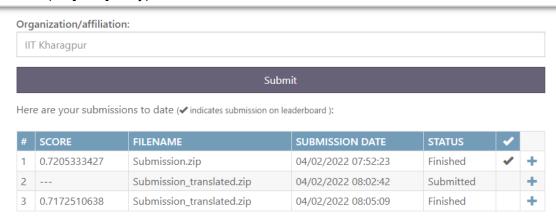
Model -> paraphrase-multilingual-mpnet-base-v2

English -> all-mpnet-base-v2

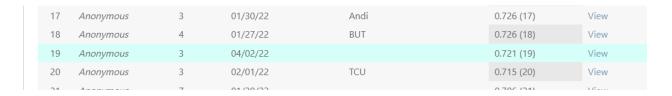
Model	Pearson Correlation Score	Link	Contributed by
ST Multilingual CS	0.5866	<u>Link</u>	Adarsh Kumar
ST Translated CS	0.6058		
ST MultiLingual + "Title"	0.6778	Link	
ST Translated + "Title" CS	0.7086		
ST MultiLingual + "Title" + CosSimLoss Finetune CS	0.7518	Link	
ST Translated + "Title" + CosSimLoss Finetune CS	0.7492	<u>Link</u>	
ST Translated + "Title" + MSE Regression	0.2864	Link	
mBERT + Translation + "Title" + CosSimLoss Finetune CS	0.6874	<u>Link</u>	Animesh Jain
XLM-R + Translation + "Title" + CosSimLoss Finetune CS	0.6357	Link	
mBERT + "Title" + CosSimLoss Finetune CS	0.6621	Link	
mBERT + Translation + "Title" + MSE Regression	0.3196	<u>Link</u>	

Note: Did tried with MLP and other regression techniques, but score didn't improve. So haven't included that in the above table. Also tried concatenating the embeddings from MultiLingual and Translated, but score again didn't show much improvement.

Codalab submission: For codalab submission, we filled the missing files as empty strings for both title and text (ie "[SEP]" only). Submission Results:



LeaderBoard (https://competitions.codalab.org/competitions/33835#results)



Link to generate csv (Will need the access to other data files (Prerun Models) (that is added as "Add data"))