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	Link	Adarsh Kumar
ST Translated CS	0.6058		
ST MultiLingual + “Title” CS	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	Link	
ST Translated + “Title” + MSE Regression	0.2864	Link	
mBERT + Translation + “Title” + CosSimLoss Finetune CS	0.6874	Link	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	Link	

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:

Organization/affiliation:

IIT Kharagpur

Submit

Here are your submissions to date (✔ indicates submission on leaderboard):

#	SCORE	FILENAME	SUBMISSION DATE	STATUS	✔	
1	0.7205333427	Submission.zip	04/02/2022 07:52:23	Finished	✔	+
2	---	Submission_translated.zip	04/02/2022 08:02:42	Submitted		+
3	0.7172510638	Submission_translated.zip	04/02/2022 08:05:09	Finished		+

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

17	*Anonymous*	3	01/30/22	Andi	0.726 (17)	View
18	*Anonymous*	4	01/27/22	BUT	0.726 (18)	View
19	*Anonymous*	3	04/02/22		0.721 (19)	View
20	*Anonymous*	3	02/01/22	TCU	0.715 (20)	View

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