

Research Project Progress Report Week [6] – SIT723

Student Name:	Oscar WU
Supervisors' Names:	Daniel Ma
Project Title:	Multi-lingual COVID19 misinformation detection and model explanation
SIT723 Target Grade:	HD
Overleaf Project Link:	https://www.overleaf.com/read/pbmgdmgghsnq
Project Folder Link:	https://github.com/wuyoscar/SIT723-Reserach-Project-A
Worklog:	137
Project Plan	
Summary of the work planned with your supervisor:	[Add details of the work you intended to complete, i.e., planned in consultation with your supervisors in the past/current week] 1. How to improve model (i.e. hyperparameter setting) 2. How to appropriately processing datasets 3. SHAP cost lot of budget and any solution to overcome this
Summary of the work done:	[Add details of the work you completed in the past/current week] If you did not manage to achieve all the planned tasks, add some explanation for the missed or incomplete tasks. 1. Implemented NLP and word-embedding 2. Advanced EDA for datasets



Next steps:	[Add immediate next steps for the project, i.e., work planned for the upcoming week] Try different subset and data processing skills to see if there are improvement on prediction accuracy. Because our datasets has attributes such as news_content, news_summary, news_titiel, news_claim. News_claim is one or two sentence to summarize this news, it performs generally better than content.
Overall project progress:	 [Add comments on the project backlog, i.e., your overall progress in the project for achieving the final results] The final results of projects aims for: Datasets contains covid19 misinformation which contains crossdomain and multiple news medium, which can be used in NLP covid19 detection research CT-BERT Model should be generally better than benchmark model. Model explanation should make public to understand what is happening inside ML black box model.