1. Group 1: The Disaster Tweets - Text Classification (CAI Shizhan, SONG Wenxin)

a. Summary of the report

Twitter has become a common platform where people report emergencies they are encountering in real-time. Consequently, more agencies are now monitoring Twitter to detect disasters. However, the line between reporting real disasters and tweeting metaphorically is subtle. This project aims at building machine learning models to distinguish whether a Tweet is describing a real disaster or not.

b. Describe the strengths of the report

Flow of data processing is clearly illustrated, and top 10 frequent locations and keywords are presented with bar charts. Two NLP models, namely RoBERTa and XLNet are used for prediction. The model structures for each of them are shown and a comparison between them is made.

c. Describe the weaknesses of the report

More details regarding the steps of processing the text data are expected. Also, more NLP models could be applied as mentioned by the group, these two methods are not much different but the location of "mask". It would be great to see a variety of models giving varying results. Also, it would be great to include hyperparameters tuning to improve the accuracy.

d. Evaluation on quality of writing (1-5): 2

Clear language is used throughout the report. I think the writing could be improved by adding the feature importance graph in the poster to better present the result of modelling. Also, the project only discussed two models, and if the objective is to distinguish whether a Tweet is describing a real disaster or not, some other models or ways to improving the accuracy (e.g. tuning the hyperparameters) should be considered. If the goal is to simply compare the performance of RoBERTa and XLNet in that set of data, it is better to state them clearly in the objective section.

e. Evaluation on presentation (1-5): 2

The presentation did include many contents that are already stated in their poster and source codes, such as procedures of data processing, details of the models selected (RoBERTa and XLNet), etc. Reasoning on each step are also provided.

f. Evaluation on creativity (1-5): 1

BERT would be one of the most common ways of delivering desirable prediction accuracy in the context of this project, which also makes it popular in NLP usage. In fact, the aggregation to RoBERTa and XLNet part should be possibly generalizable to many other NLP problems in addition to the one proposed in this project. While there is a lack of specificity towards the problem, and both models are commonly used, some useful data insights are expected in the conclusion section in the project.

g. **Confidence on your assessment (1-3):** 3

I have carefully read the paper and checked the results.