**Progress Report**

**9 June 2024**

1. Data is fully in hand. ✅
2. EDA is still under progress.
3. The modelling process is not started yet.
4. I have challenges in:
   1. Selecting the suitable data columns for NLP Sentiment Analysis modelling, as well as which column as the target. How to determine the Correlation Coefficient threshold to select the dataset features?
   2. Data cleaning. The data index is large (240k+ rows), still taking time to filter the null values.
5. I need to get at least a modelling process done this week, yet I am still taking time on EDA process.

**23 June 2024**

1. I changed to a new dataset as the previous dataset was not suitable for NLP use. Dataset is fully in hand. ✅
2. EDA is completed. ✅
3. Modelling process is completed. ✅
4. Model evaluation is completed. ✅
5. I have limitations in:
   1. Performing non-English text translation due to my computer hardware limitation, it would take around 7-8 hours to complete the text translation for ONE column, and I have TWO columns’ data text need to be translated.

**28 June 2024**

1. I changed to a new dataset as the previous dataset was not suitable for NLP use. Dataset is fully in hand. ✅
2. EDA is completed. ✅
3. Modelling process is completed. ✅
4. Model evaluation is completed. ✅
5. My limitations:
   1. Performing non-English text translation due to my computer hardware limitation, it would take around 7-8 hours to complete the text translation for ONE column, and I have TWO columns’ data text need to be translated. Hence, I remove all the columns with non-English text instead going for text translation due to time constraint. It is recommended to use a computer with appropriate hardware specification to handle a large data set.
6. Project is completed. ✅