**Assignment Response**

**Results Summary:**

1. The Baseline model performed better than the SOTA model
2. The above observation is mainly impacted by the choice of data preprocessing.
3. For the baseline model, it was observed that most of the predictions ranking are within 1 or 5 ranting, meaning either best or worst based on inference from confusion matrix.

**Attachment:**

The following file are attached:

**Data file:**

* Preprocessed training data (x\_train and y\_train), development data (x\_dev and y\_dev) and testing dataset (x\_testing).
* Prediction Files
  + Baseline Model
  + NN Model

**Code:**

* Base Model: contains the following files:
  + Bae ModelV2
  + Development\_Test\_Evaluationv2
* NN Model

**Challenges:**

* Could not get GPU for data preprocessing using better methods
* I tried creating a code for creating a summary of the review but had to abort it because it requires some amount of time. The review summary could have been a better approach than the vocabulary approach.

**References:**

* Ahilan Srivishnumohan (2019) ‘Sentiment Analysis for Hotel Reviews’, Toward Data Science, [online] Available at: https://towardsdatascience.com/sentiment-analysis-for-hotel-reviews-3fa0c287d82e