**Assignment Question**

**Data description:**

The assignment is a sentiment classification task for hotel reviews. There are 5 ratings that are the labels to be predicted based on the text in the reviews. You have been given the ratings for the train and dev sets. The ratings on the test set have been withheld.

Data Download Path: <https://chataai.sharepoint.com/:f:/g/EtI-zi28qBJKjafd6BVpkTgBNM906KqOA3zfZU4fXXMGeA?e=fMEKGx>

**Task Description:**

Please build 2 models for this text classification task –

1. A Baseline model and 2. A State of the art (SOTA) model and compare them.

**Artifacts to be submitted:**

1. Please submit your model evaluation results for the 2 models for both dev and test sets.

The predictions for each row must be shown.

1. Code for the models (train, evaluation, inference) and trained model files. We must be able to run an inference script to test the performance of the models on other held-out test sets.
2. A writeup on your solution/process – analysis, research, data exploration, model selection, evaluation measures, model comparison etc. Please include references.

**Scoring:**

The assignment will be graded based on the following criteria and in this priority order.

1. Model accuracy.
2. Clean, Original, Readable Code. (Use of off-the-shelf libraries is discouraged)
3. Research and Justification.
4. Analysis on data, model, performance.

**Submission**

You have 7 days to complete and submit the assignment, starting from the day you receive it. Please upload the artifacts to a Github Repo and share with us, in a response email.

Thank you.

We look forward to working with you.