Q1. Form a unigram language model (25 marks)

Implement unsmoothed and smoothed unigram language models in python using any available techniques and report the output.

Q2. Form a bigram language model (25 marks)

Implement unsmoothed and smoothed bigram language models in python using any available techniques and report the output. (12 marks)

Q3. Work on sentence probabilities (25 marks)

Using the *smoothed* model values, carry out the following tasks.

1. Compute manually the sentences probabilities using the unigram model. (5 marks)
2. Compute manually the sentences probabilities using the bigram model. (5 marks)
3. Justify which language model is more suitable to calculate the sentence probabilities. (5 marks)
4. Implement and report the respective sentence probabilities in python using both unigram and bigram language models. (10 marks)

Q4. Supervised Text Classification (25 marks)

Use the **Musical\_Instruments\_Reviews.csv** data set available in the **Group Assignment Data** to perform the following tasks.

1. Export the resulting data set along with the predicted sentiments to a .csv format using python and report the relevant code used for this operation. *(Note: The resulting data set must be submitted via the given submission link in MOODLE).* (5 marks)
2. Use the data set exported in the Q4.2 to build a supervised sentiment classification model using *Naïve Bayes Classifier* from NLTK and report the following model performance measures:
3. Accuracy
4. Precision
5. Recall
6. F1 Score

The python codes must be neat with clear output. Provide relevant comments in the code to explain the purpose of the code snippet.

Deliverable:

1. Complete & running Python code.
2. Output stating the above FOUR (04) performance measures.

(20 marks)