Assignment 1 - CT5120
Chin Zhe Jing 22221970

1. **Define suggestion mining in your own words.**
    a. In a large volume of unstructured text, there might be some suggestions embedded in the context. Instead of using human eyes to scan across the big amount of text, suggestion mining helps to extract the exact sentences that is giving suggestion and provide useful information for decision making.

2. **Explain a use case where suggestion mining could be useful.**
    a. A seller could use information extracted from suggestion mining to understand the needs of their customers, thus making improvement on their products. For example, a start-up skincare company could mine suggestions from Instagram comments on their post and research / release a new product that has highest voting.

3. **Give any two challenges involved in the suggestion classification task with short explanation.**
    a. Long and complex sentences:
        i. Suggestion might only expressed in one part of a very long sentence. Hence, annotator tend to label wrongly.
    b. Task formalization and data annotation:
        i. Human understand suggestions differently, hence providing difficulties in labelling suggestion sentences and bias tend to occur.

4. **Explain implicit and explicit suggestions in your own words along with an example for each.**
    a. Implicit suggestion:
        i. Suggestion is indirectly stated in a sentence.
            1. Example: “X hotel is night-owl friendly with breakfast provided until 12pm.”
    b. Explicit suggestion:
        i. Suggestion is written directly in a sentence often with specific keywords for recommend, request, suggest, need / necessity.
            1. Example: “I recommend you sell sunscreen with SPF 50 +++.”

5. **Is the following sentence a suggestion: “I would not travel to the USA during the pandemic?” Why or why not?**
    a. In my opinion, the above sentence is not a suggestion as it did not specify the actual advise, hence it might just be a personal travel plan.

6. **Give an example where more context for a sentence could possibly turn a non-suggestion into a suggestion?**
    a. For example, adding reason for the below sentence could possibly turn non-suggestion into a suggestion:
        i. I would not travel to the USA during the pandemic as the US government did not give out free vaccination for travellers.

1. It could possibly become a suggestion for US tourism department to promote the country and attract more travellers.

7. **For one crowdsourcing platform, state the advantages and disadvantages of such a platform.**
    a. Crowdflower
        i. Advantages:
            1. Tasks assigned based on experience level with pass score therefore researchers could have a more complete control on the quality.
            2. Able to reduce the amount of annotations needed from the experts
            3. Cheaper cost
        ii. Disadvantages:
            1. Guideline and training period needed for layman
            2. Ambiguous sentence hard to be annotated

8. **How is inter-annotator agreement used for the suggestion mining task?**
    a. Inter-annotator agreement is having two expert annotators labelling a set of sentences and calculate the score of the total sentences they both agreed to the same decision. Such score determines the quality of the annotation guidelines and the reliability of the annotated set.

9. **How will you evaluate a text classification model on a benchmark suggestion classification dataset?**
    a. I will use cross-validation as it randomize and train the dataset for multiple sets then compare the results with the labelled tags. After that, performance metrics including confusion matrix, accuracy, precision, recall, and f1 score can be generated to numerate the performance of the model.

10. **Suggest one other text classification task similar to suggestion mining. Does it need an annotated (supervised) dataset?**
    a. Sentiment analysis on tweets, classifying a text to either 3 classes (positive, neutral , or negative). Sentiment analysis can be solved with supervised or unsupervised algorithm. For supervised algorithm, we need to provide sentiment annotation dataset with sentences and their tag with either positive / negative (binary classification) or neutral (for multi-class classification) .