**Assignment 1**

1. **Define suggestion mining in your own words.**

**Answer**: Suggestion mining or Opinion mining is a subtype of data mining where you extract sentences that resemble suggestions from an unstructured corpus. Collecting suggestions is an important step in the decision-making process that any business undertakes, hence suggestion mining can be used while planning improvements or modifications to any system.

1. **Explain a use case where suggestion mining could be useful.**

**Answer**: Suggestion mining would be extremely useful for CRM (Customer Relationship Management) systems. In CRM systems, customer data is stored in a structured format which includes customer feedback and reviews. After aggregating customer reviews, suggestion mining could be used on these reviews to extract customer opinions which can later be used to improve the business intelligence of the current application.

1. **Give any two challenges involved in the suggestion classification task with short explanation.**

**Answer**: 1. Data annotation: There is a limited labelled dataset in this task, hence it is difficult to formulate any task related to suggestion mining and data needs to be first labelled by understanding sentence structuring or by crowdsourcing annotation task as cited by the authors of the research paper *Open domain suggestion mining.*

2. Context dependency: Sometimes, to identify a sentence you need some additional context. For example, *Epson ink cartridge is a good addon, relatively cheap.* This sentence alone cannot be considered a suggestion. Suppose we are told that the above sentence is from a printer review. Given the following context, we can assume that the review suggested that the ink cartridge is a good addon to buy along with the printer.

1. **Explain implicit and explicit suggestions in your own words along with an example for each.**

**Answer**: 1. Explicit suggestion: They are sentences that are always returned by the suggestion function as a suggestion with or without any additional context. These are also referred to as direct or conventional suggestions as they are rather easy to figure out.

For example: *If you wish to visit this restaurant, be sure to book early as tables are not available later in the evening.*

1. Implicit suggestion: These are sentences that are returned by the suggestion function as non-suggestion when the additional context is not provided. Predominantly, these sentences alone are inadequate to conclude whether they are suggestions or not, but additional context helps the function to identify them better.

For example: *There is a bus stop about 400 meters away, good connectivity.* This sentence alone is not enough to label, but with the context that the sentence is in a restaurant review. It becomes clear that the sentence is a suggestion.

1. **Is the following sentence a suggestion: “I would not travel to the USA during the pandemic?” Why or why not?**

**Answer:** No, the sentence is not a suggestion, the sentence is more of an assertion. Where the subject is asserting that he will not be traveling to the USA during the pandemic. However, it can be inferred that the conditions on the subject can be valid for others as well. In that case, it is an indirect suggestion.

1. **Give an example where more context for a sentence could possibly turn a non-suggestion into a suggestion?**

**Answer:** *MagSafe wireless charger is a good option to purchase.* This sentence alone is not a suggestion. But with the context that this sentence is taken from a smartphone review, It becomes clear that it is a suggestion to buy a certain charger that complements the smartphone mentioned in the review.

1. **For one crowdsourcing platform, state the advantages and disadvantages of such a platform.**

**Answer:**

In the research paper, open domain suggestion mining authors hire annotators from Crowdflower a crowdsourcing platform.

Advantages:

1. You can compare annotation results from various regions and thus compare labeling from different perspectives.
2. If you don’t have enough annotated data crowdsourcing can be helpful as you can generate the dataset.

Disadvantages:

1. In the study, the authors observed that workers on crowdsourcing platforms sometimes could not understand guidelines or even ignored them**.**
2. Layman workers were not best suited for carrying out annotation tasks and trained volunteers were slow and expensive to hire.
3. **How is an inter-annotator agreement used for the suggestion mining task?**

**Answer:** Inter-annotator agreement uses a method called Cohen’s kappa. It is used to compare two distinct observations. Annotators of the observations should be ordinal type. In suggestion mining, we are annotating our samples as non-suggestion and suggestions. The inter-annotator agreement is used to compare two or more sets of annotators for a sentence and give a similarity value of the sets.

1. **How will you evaluate a text classification model on a benchmark suggestion classification dataset?**

**Answer**: A benchmark suggestion classification dataset would have annotated data i.e. sentences would be labelled as suggestion or non-suggestion. In this dataset, each sentence could be converted to a vector using a bag of words or any other vectorization model with a corresponding label (suggestion, non-suggestion). We have a set of features (vector) and a dependent variable(label). We can use classification algorithms like Logistic Regression, and Support Vector Machine to model a hypothesis based on the dataset. This model could be tested on unseen data to validate the model accuracy and later be used for text classification tasks.

1. **Suggest one other text classification task similar to suggestion mining. Does it need an annotated (supervised) dataset?**

**Answer:** We can mine for rhetoric questions in the dataset in the same vein as suggestion mining. Rhetorical questions are questions that are asked to create a dramatic point instead of looking for an answer. For example, *what’s not to like?* A question without an answer is a rhetorical question. We do not need an annotated dataset. We can mine for rhetoric questions from a corpus by analysing sentence structuring.