

Assignment-1 Frequency Analysis and Bag of Words

Tag: MDS555-2023-Assignment-1

Objective

- Learn the keywords of NLP
- Understand basic of NLP
- Explore the Bag of words for basic analysis of text

Tasks

- 1) Task 1: Dataset Preparation: Prepare the Nepali news dataset (*hint: you can obtain text from news websites, at least 20 different news of 2/3 different categories*).
Host the dataset in the public git repository.
In your notebook data should be downloaded from git or some other public places.
No additional step should be done to get the notebook working.
- 2) Task 2.1: Frequency Analysis : Perform the frequency analysis on the text collected
- 3) Task 2.2: Filter Stop words :Improve Performance analysis by filtering stop words (you can also develop rule based)
- 4) Task 3: BoW: Prepare Bag of Words (BoW) from the dataset
- 5) Task 4: Classification: Classify the news based on Keywords and BoW you computed in Task 3.1

Deliverable

The deliverable should be in ipython notebook format. The template has been uploaded in the link below.

Submission Template: [\[View in Colab\]](#)

<https://colab.research.google.com/drive/1YfWcNSkujPOmJwLSi1aF5YoxMhZ7EgsV?usp=sharing>