

CSE 4221

NLP Assignment

Deadline: 18/03/2021 (11.59 PM)

Steps:

1. Choose a genre of your own. Example Bangladesh railway, Bangladesh cricket, COVID19, and so on.
2. Develop a corpus of at least 100 sentences. You can increase your corpus size based on your genre. This is preferable.
3. Take an input sentence.
Example: Sundarban express is traveling from Dhaka?
4. Find out the correct words to complete the sentence.
5. Apply unigram, bigram, and trigram method or n-gram as a whole.
6. Check the zeros.
7. Apply smoothing techniques based on your problem. Example add-1, add-k, stupid backoff, and Kneser-Ney.
8. Generate probability table.
9. Make a discussion of your result.
10. Do not copy each other. If the corpus is copied, you will get zero, for sure!!!.

Submission rule:

1. Write your report in MS word/ PDF or hand written scanned/image/photo. **I need to read only.**
2. Save as your roll (Example: 1607048.pdf)
3. Upload your assignment corresponding google drive link. CR will provide the drive link later. (CR (1007048) create a link please!!! and send)