CS3216/AI5203
Assignment 2- NLP

Deadline: 4th March, 2024 **Time**: 11:59 PM

Total Marks: 100

Instructions: -

- Please don't copy from the internet or any other student.
- Plagiarism will thoroughly be checked for all submissions.
    - Refer to the policy on Academic Unfair means.
- You may use any libraries such as spacy, nltk, regex libraries.
- Allowed Programming Language: Python
- It is recommended to use Jupyter Notebook/Google colab
- **Submission link: https://forms.office.com/r/dEBaC28Q7X**
- The format of the submission file should be: **RollNo-A2.ipynb**
    - The python notebook file should contain:
      - The code in a step-by-step format with the results of each stage
      - Readable comments for each stage
      - Short 1-2 lines analysis/observations of the results

Problem Statement: Implement Word2vec from scratch.

1. Train a neural network to generate word embeddings in the form of a weight matrix.
2. Embedding layers can be trained to generate custom embeddings in popular neural network libraries like TensorFlow or PyTorch.
3. You are free to train on large datasets such as Wikipedia.
4. Test the semantics, embedding vectors to see relationships such as "king - man + woman = queen."

---------All the best--------