



Discipline of
Computer Science & Engineering

Assignment 1 - Executive M.Tech(AI)
Course: NLP

Due Date: Open

Instructions

1. Assignment submissions will be accepted only via Google Classroom. Submissions through email or any other methods will NOT be accepted. Please join Google Classroom using the following link: <https://classroom.google.com/c/NzUyOTIwOTIxOTU2?cjc=y66aldcx>
 2. This is a graded assignment (10 points). Penalties may be applied to those who do not submit the assignment before due date.
 3. The submission deadline is one week before end-term exam. Please submit a single .pdf file using the nomenclature 'NLP<AssignmentNumber>_<EnrollmentNumber>.pdf,' for example, 'NLP1.2022RCS2021.'
-

1. You are provided with a dataset containing text data for sentiment analysis (uploaded with the assignment on Google Classroom). The dataset includes various financial sentences along with their corresponding sentiment labels. Your task is to perform the following tasks:
 1. Clean and preprocess the given data. This includes tasks such as removing stopwords, handling punctuation, and tokenizing the text. Ensure that the data is in a suitable format for further analysis.
 2. Implement Continuous Bag of Words (CBOW) and Skip-gram (Skip-gram) word embedding models on the preprocessed data. You can use existing libraries or implement the models from scratch.
 3. Implement an LLM model to generate embeddings from the preprocessed data. You may use existing pre-trained LLMs.
 4. Develop at least three different ML models of your choice for sentiment analysis. You are free to choose the architectures, but they should be suitable for the sentiment classification task. Train the models using the preprocessed data.
 5. Evaluate the performance of your ML models using the different generated word embeddings. Compare and analyze the results. Consider metrics such as accuracy, precision, recall, and F1 score. Discuss any differences observed between the various types of word embeddings.

Be sure to include code snippets, visualizations, and a comprehensive explanation of your methodology and discoveries in your submission as a single PDF file. Please avoid submitting the .ipynb file directly; instead, convert it to a PDF format. The objective is to demonstrate your proficiency in data preprocessing, word embeddings, neural network modeling, and your insights into the effects of various word embedding techniques on sentiment analysis performance.