भारतीय प्रौद्योगिकी संस्थान जम्मू
INDIAN INSTITUTE OF TECHNOLOGY JAMMU

Discipline of
**Computer Science & Engineering**

---

**Assignment 1 - Executive M.Tech(AI)**
**Course: NLP** **Due Date: 21/09/2024**

---

**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://shorturl.at/IAqx2
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 21/09/2024. 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 assignment on google classroom). The dataset includes various financial sentences along with their corresponding sentiment labels. Your task is to perform the following tasks:
   - 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.
   - Implement Continuous Bag of Words (CBOW) and Skip-gram (SKI gram) word embedding models on the preprocessed data. You can use existing libraries or implement the models from scratch.
   - Develop a Neural Network (NN) model of your choice for sentiment analysis. You are free to choose the architecture, but it should be suitable for sentiment classification task. Train the model using the prepossessed data.
   - Evaluate the performance of your NN model using both CBOW and Skip-gram word embeddings. Compare and analyze the results. Consider metrics such as accuracy, precision, recall, and F1 score. Discuss any differences observed between the two 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.