**Sentiment Analysis using NLP**

**Instructions:**

In this assignment, you will demonstrate your understanding of sentiment analysis using Natural Language Processing (NLP) techniques.

Dataset: Select a dataset suitable for sentiment analysis tasks. You can use publicly available datasets such as IMDb movie reviews, Twitter sentiment analysis, or any other dataset that includes text data labelled with sentiment (positive, negative, or neutral).

**Steps to follow:**

a. Dataset preprocessing: Preprocess the dataset by cleaning the text, removing any noise or irrelevant information, and performing any necessary text normalization techniques such as tokenization, stemming, or lemmatization.

b. Split the dataset: Split the dataset into training and testing sets. Use an appropriate ratio (e.g., 70% training, 30% testing).

c. Feature extraction: Convert the textual data into numerical features that can be used by machine learning models. Use techniques such as bag-of-words, TF-IDF (Term Frequency-Inverse Document Frequency), or word embeddings (e.g., Word2Vec, GloVe).

d. Model selection and training: Choose a suitable machine learning or deep learning model for sentiment analysis, such as Naive Bayes, Support Vector Machines (SVM), Recurrent Neural Networks (RNN), or Transformers (e.g., BERT). Train the selected model on the training dataset.

e. Model evaluation: Evaluate the trained model on the testing dataset by calculating relevant evaluation metrics such as accuracy, precision, recall, and F1 score. Assess the model's performance in predicting sentiment labels.

f. Fine-tuning and optimization: Experiment with hyperparameter tuning and model optimization techniques to improve the performance of the sentiment analysis model. Explore techniques such as grid search, cross-validation, or ensemble methods to enhance the model's effectiveness.

g. Prediction and analysis: Use the trained model to predict the sentiment labels of new, unseen text samples. Analyse the model's predictions and assess its strengths and limitations in handling different sentiment expressions and contexts.

h. Document your findings: Provide a detailed explanation of the steps taken, the observations made, and the results. Include visualizations, if applicable, to support your analysis.

Introduction: Submit your assignment as a pdf, and video file that includes the following sections

Dataset: Describe the dataset used, including its source, size, and sentiment label distribution.

Methodology: Describe the preprocessing techniques, feature extraction methods, model selection, and training strategies employed.

Results: Present the evaluation metrics, including accuracy and other relevant performance indicators. Discuss the model's strengths and weaknesses.