Logo

Description automatically generated **FACULTY OF COMPUTER SCIENCE AND ENGINEERING**

**Time: 1 hours AI 361 Lab Marks: 80**

**Instructor: Muhammad Muneeb Baig Lab #03**

**Task 01:** Find out the number of words in the entire corpus and also the total number of unique words(vocabulary) using just python.

**Task 02:** Apply bag words and find the vocabulary also find the times each word has occurred.

**Task 03**: <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews?resource=download>

* This data consists of two columns. - review - sentiment
* Reviews are the statements given by users after watching the movie.
* sentiment feature tells whether the given review is positive or negative.

**Exercise-1**

1. using sklearn pipeline module create a classification pipeline to classify the movie review's positive or negative.

**Note:**

* use CountVectorizer for pre-processing the text.
* use **Random Forest** as the classifier with estimators as 50 and criterion as entropy.
* print the classification report.

**References**:

* <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
* <https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html>

**Exercise-2**

1. using sklearn pipeline module create a classification pipeline to classify the movie review's positive or negative..

**Note:**

* use CountVectorizer for pre-processing the text.
* use **KNN** as the classifier with n\_neighbors of 10 and metric as 'euclidean'.
* print the classification report.

**References**:

* <https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html>
* <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

**Exercise-3**

1. using sklearn pipeline module create a classification pipeline to classify the movie review's positive or negative..

**Note:**

* use CountVectorizer for pre-processing the text.
* use **Multinomial Naive Bayes** as the classifier.
* print the classification report.

**References**:

* <https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html>
* <https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html>