Information Retrieval (CS4051)
Programming Assignment No. 3
Spring 2024

Submission Date: May 06 2024

Assignment Objective

This assignment is related to text classification and clustering. There are two parts of this assignment.

Part No. 1 Text Classification

You need to design a text classification algorithm using vector space data prepared in A2. Using a k-NN classifier you need to classify the test data that will be provided later. You can use ML library for this part. The earlier provided dataset is coming from 5 classes, here are the label data for each provided document.

Class Name: "Explainable Artificial Intelligence"
Documents: 1,2,3, and 7

Class Name: "Heart Failure"
Documents: 8,9, and 11

Class Name: "Time Series Forecasting"
Documents: 12,13,14,15, and 16

Class Name: "Transformer Model"
Documents: 17,18, and 21

Class Name: "Feature Selection"
Documents: 22,23,24,25, and 26

After the model building and classification you need to obtained the following score for evaluation of your task. Precision, Recall, F1 and Accuracy is used for the evaluation.

Part No. 2 Text Clustering

In this part you will be using the same dataset and for this part you will be doing text clustering using k-Means algorithm. The same vector space can be utilized for performing document vector clustering. Your implementation will be using an input parameter k for obtaining the final clustering arrangement. The clustering results will be evaluated by using following evaluation metrics: Purity, Silhouette Score and Random Index.

Evaluation/ Grading Criteria

The grading will be done as per the implementations, performance and evaluation metrics

Grading Criteria:
Model Implementation (2.5 marks)
Model Timing (2.5 marks)
Evaluation metrics (2.5 marks)
Code and Comments (2.5 mark)
Bonus: GUI (1 mark for making the GUI 1 mark for Good Looking GUI)

<The End>