SVKM's NMIMS

Mukesh Patel School of Technology Management & Engineering

Department of Artificial Intelligence

Program: B. Tech/MBA Tech Semester: V

Course: Natural Language Processing

Self Practicum

Name: Pranav Dhadwal, Aditya Kothari

Roll Numbers: I016, I028

(PART A: TO BE REFERRED BY STUDENTS)

Aim:

The aim of this project is to introduce students to the practical implementation of the K-Nearest Neighbors algorithm for text classification and to familiarize them with the essential concepts and techniques involved in document categorization tasks.

Prerequisites:

Basic understanding of Python programming.

Familiarity with fundamental concepts of machine learning and text preprocessing.

Knowledge of the K-Nearest Neighbors algorithm and its application in classification tasks.

Outcome:

Upon completion of this project, students will:

Understand the implementation of the KNN algorithm for text classification.

Gain practical experience in data preprocessing and feature engineering for text data.

Learn how to evaluate and fine-tune a KNN model for optimal performance in document classification tasks.

Theory:

K-Nearest Neighbors (KNN) is a simple and effective algorithm used for both classification and regression tasks. In the context of text classification, KNN works by identifying the k closest training examples in the feature space and predicting the label of the test sample based on the majority vote of its nearest neighbors.

Tasks to be Completed:

Data collection and preprocessing.

Implementation of the KNN algorithm for text classification.

Training and evaluation of the KNN model.

Hyperparameter tuning for optimizing the KNN model's performance.

Analysis of misclassified documents and potential reasons for misclassifications.

PART B

(PART B: TO BE COMPLETED BY STUDENTS)

Code :
https://colab.research.google.com/drive/10SDeGNjcEuPNLxJCfsonWTUuDFQC5ZMO?usp=sharing

Conclusion:

In conclusion, this project provided valuable insights into the application of the K-Nearest Neighbors algorithm in document classification tasks. Through the implementation and evaluation of the KNN model, students gained practical experience in handling text data and learned the importance of data preprocessing and hyperparameter tuning in enhancing the performance of machine learning models.

Result:

The KNN model achieved an accuracy of [insert accuracy score] on the test dataset after hyperparameter tuning. The model performed well in categorizing the text documents into predefined classes, demonstrating the effectiveness of the KNN algorithm in text classification tasks.

References:

Scikit-learn Documentation. Available at - https://scikit-learn.org/stable/documentation.html