Lab Report on

K Nearest Neighbor

Submitted to

Department of Computer Science and Engineering

Nepal Engineering College

in Partial Fulfillment of the

Requirements for the Degree of B.E. in Computer

Submitted By

Dipson Thapa (020-313)

Date: *28/07/2024*

K Nearest Neighbors

Introduction:

K-Nearest Neighbors (KNN) is a simple, yet powerful supervised machine learning algorithm used for both classification and regression tasks. It works by finding the k nearest data points in the training set to a given input point and making predictions based on the majority class (for classification) or average value (for regression) of these neighbors. The algorithm is non-parametric, meaning it makes no assumptions about the underlying data distribution. KNN is widely used in pattern recognition, data mining, and intrusion detection due to its simplicity and effectiveness.

Dataset:

The dataset used for this K-Nearest Neighbors analysis is the "Iris" dataset, which contains measurements of different features of iris flowers from three species. You can access it here.

Link: https://www.kaggle.com/datasets/rajakali/diabetesknn

Notebook name: K_Nearest_Neighbor_020313.ipynb

Libraries Used:

1. **Pandas**:
    - Utilized for data manipulation and analysis. It offers data structures like DataFrames that are ideal for handling tabular data.
2. **Pathlib**:
    - A standard library for handling filesystem paths in a more readable and efficient way.
3. **numpy**:
    - A fundamental package for numerical computations in Python, providing support for arrays and matrices along with a collection of mathematical functions to operate on these data structures.
4. **Scikit-Learn (sklearn):**
    - **sklearn.model_selection**:
        - **train_test_split**: Utility for splitting the dataset into training and testing sets.

- **sklearn.preprocessing**:
    - **StandardScaler**: Used for feature scaling to ensure that each feature contributes equally to the distance calculations during clustering.
- **sklearn.neighbors**:
    - **KNeighborsClassifier**: A KNN model from the scikit-learn library used for classification tasks.
- **sklearn.metrics**:
    - **accuracy_score** and **confusion_matrix**: Metrics for evaluating the performance of the KNN model.