**ASSIGNMENT – 1**

**A Machine Learning project with Python**

**Objective:**

This paper aims to evaluate the effectiveness of the K-Nearest Neighbours (KNN) classification method for the targeted study. The classifier's performance is assessed while adjusting k, which entails comprehending how selecting the right k affects the MCD. A heatmap is used to depict the accuracy across a range of k parameters from 1 to 10.

**Methodology:**

**1. Data Generation:**

A synthetic dataset was made using the sklearn's make\_blobs function, which allowed the output to be modified by changing the random state variable to five times the dataset's feature count. the datasets module.

**Dataset Characteristics:**

The dataset includes 310 samples which are grouped into three clusters with the points (2, 4), (6, 6) and (9, 1).

**2. Data Splitting:**

After that, the full set of data was separated into the target vector Y and the features matrix. The train\_test\_split function was then used to divide the entire database into a test set and a training set, with the test set making up 20% of the data.

Training data was used to build the KNN model, while testing data was utilised to assess the model's effectiveness.

**3. KNN Model Implementation:**

The KNN classifier was initially set with value of k = 6 neighbors for testing purposes.

The training data was used to build and develop the model, while the test data was used to test the model.

In this instance, the accuracy\_score function from the sklearn.metrics package was used to assess how accurate the models' predictions were.

**4. Accuracy Testing for Different k Values:Accuracy Testing for Different k Values:**

In order to check the impact of the change of k, the accuracy of the classifier was computed by taking k as 1, 2, …, 10.

A heatmap was created in order to represent all these accuracy scores based on different k values.

**Results:**

**Classifier Accuracy with k=6:**

When k=6, the KNN classifier's classification performance was mediocre, with a test accuracy of 0.98. This suggests that the classification accuracy of test samples using six nearest neighbours was 98%.

Accuracy at Various k Values: Accuracy at Various k Values:

Throughout the experiments, the accuracy level remained constant at roughly 0.98 for all k values between 1 and 10. This demonstrates that the model functions well regardless of the number of neighbours utilised and that the dataset is quite tiny.

The accuracy ratings varied slightly, but they were nearly always high regardless of the value of k, as the heatmap below shows. This supports the notion that changing of k has little effect on the model's performance in the future because the dataset is well-organised. A blue bar graph with white text

Description automatically generated

**Discussion:**

High Accuracy Across All k Values: Nevertheless, it was found that the KNN classifier's efficiency barely changed, averaging almost 98% across all "k" values. This indicates that the dataset's clusters can be distinguished, which makes it simpler for the classifier to generate accurate predictions for the input values.

**Model Stability:**

Regarding the k value, it is commonly assumed that a number that is too low will cause the model to be overfit, while a value that is too high will cause the model to be underfit. The steady accuracy in this instance, however, raises some concerns because it suggests that the dataset does not present a challenging classification assignment for the model.

**Visualization:**

Heatmaps are preferred because they provide a clear and concise picture of the classifier's performance across a range of "k" values. Since the entire silhouette of the heatmap is dark blue, they were able to get a high degree of accuracy.

**Conclusion:**

Applying the KNN classifier to this artificial data set yielded an accuracy of 0.98 for all values of k = 1–10. This indicates that the data items are clean for the classification process and that the dataset utilised in this work is simple. More complex test data sets with intersecting classes can be used in subsequent experiments to more thoroughly examine the impact of variations in k.