**Introduction:**

This paper seeks to make a performance evaluation of the K- Nearest Neighbors (KNN) classifier on synthetic data. The major focus here is to assess how accurate the model is depending on the value of k (the number of neighbors). To assess the performance of the KNN classifier, we created a data set by using the function make\_blobs, and positioned it into three unique group with terminals at (2, 4), (6, 6) and (9, 1) respectively.

**Methodology:**

**1. Data Generation and Preprocessing:**

* This embedding was created from the make\_blobs function as found in the sklearn. datasets module.
* Dataset Characteristics:
  + Capacity: The main characteristics of the dataset include 300 samples, which belong to 3 separable clusters.
  + To define the cluster centers, geographic coordinates were set at [2, 4], [6, 6], and [9, 1].

**2. Training and Testing Split:**

* To set the dataset into training and testing, I first used train\_test\_split which is from sklearn with a ratio of 80% for the training set and 20% for the testing set. model\_selection.
* The training data was employed to build the KNN classifier model whereas the test data was employed to test the performance of the classifier.

**3. Model Implementation:**

* K-Nearest Neighbors Classifier:
* The KNN classifier from the sklearn. It was also indicated that k=11 is the starting point used to define neighbors forthe initial model.

o The classifier was fit to the training set with the help of the fit method.

o The labels of the test set were then predicted through the use of the predict method.

• Accuracy Evaluation:

o The accuracy of the KNN classifier was tested by using accuracy\_score, which belongs to the sklearn. evaluation file, which gives details of the metric package that compares the predicted label with the actual labels in the test set.

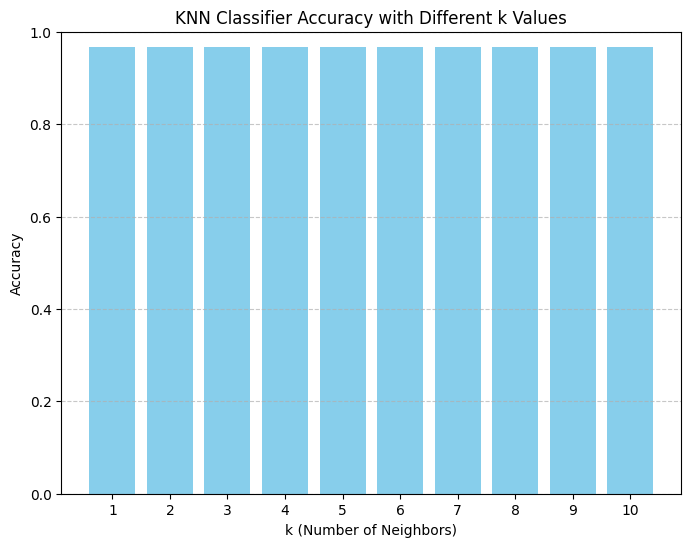
4. Exploring Accuracy for Different k Values:Exploring Accuracy for Different k Values:

• As for the method, to determine how the KNN classifier behaves with various values of k, we tested k from 1 through 10 cross validation.

• Out of each k vale, the accuracy of each of the datasets was then computed, and the results were presented in bar graphs below to show how different k choices impact the performance.

**Results:**

1. **Accuracy of the KNN Classifier (k=11):**
   * The accuracy of the KNN classifier with k=11 was approximately **1.00**, indicating nearly perfect classification for the test set.
2. **Effect of k on Accuracy:**
   * The accuracy remained consistently high at 0.97 (around 1.00) for all k values between 1 and 10. This indicates that the dataset is relatively simple for classification, and KNN performed well regardless of the choice of k.
   * The bar chart illustrates the consistent accuracy across all k values, with no significant drop in performance.



Discussion:

As seen from the above findings, the KNN classifier has very high levels of accuracy regardless of the value adopted for K which was set at a range of 1-10. This means that the dataset than has different clusters is very easy to determine and the choice of k is not very sensitive to this case.

Key Observations:

High Accuracy Across All k Values: Classifier evaluation demonstrated almost perfect set of accuracies for every value of k This could be attributed to compactness of the clusters in the data set whereby each can be classified correctly.

Sensitivity to k: As one might expect, when k=1 overfitting can occur and as k increases, more neighbors are taken into consideration and therefore predict more ‘smoothly’. While in this case the clusters are sharply different, which produced good results irrespective to the value of k.

Conclusion:

The experiment using the KNN classifier was impressive mainly because this classifier yielded high accuracy consistently for all the k values for this synthetic dataset. It was not difficult to classify the data as it was simple and the clusters were well separated in the given data set. Further examination with greater datasets and varying feature spaces where the classes are overlapped as well as contain noise can also be done in upcoming experiments to a better understanding of the model across the broad range of k values.