Summary Report – KNN Classification

Git Hub Link: https://github.com/Gaddam174/agaddam4\_machine-learning-.git

# 1. Objective

This study aims to assess the K-Nearest Neighbors (KNN) classification algorithm's performance utilizing a range of K values. Finding the ideal value of K for the greatest classification performance and evaluating the model's accuracy on training and testing datasets are the main goals.

# 2. Dataset Description

This experiment's dataset is a synthetic dataset that can be used to illustrate classification challenges. It is suitable for assessing the KNN algorithm's performance in handling classification problems because it has several classes with distinct borders.

# 3. Methodology

Training and testing sets of the dataset were separated in order to assess generalization performance. There were different numbers of neighbors (K = 3, 5, 7, and 10) when the KNN method was used. Both the training and testing datasets correctness was noted for every setup. Choosing the model with the highest testing accuracy allowed for the determination of the ideal K value.

# 4. Results

|  |  |  |
| --- | --- | --- |
| K Value | Training Accuracy | Testing Accuracy |
| 3 | 1.00 | 1.00 |
| 5 | 1.00 | 1.00 |
| 7 | 1.00 | 1.00 |
| 10 | 1.00 | 1.00 |

Best K found: 3

# 5. Visualizations

The distribution of the dataset and the KNN classifier's decision boundaries are depicted in the following visualizations:   
A scatter plot that displays the class distribution.  
Plots showing decision boundaries for K values of 3, 5, 7, and 10.

.

A diagram of different colored dots

AI-generated content may be incorrect.A diagram of a different colored scheme

AI-generated content may be incorrect.

# 6. Analysis and Observations

- For both training and testing, the KNN classifier showed flawless accuracy (100%) across all K values.   
-This shows that the dataset is clear and evenly distributed, which facilitates classification.   
Despite the equal performance of all K values, the best K found was 3, which had fewer neighbors but still achieved the same accuracy.   
-A The results demonstrate the robustness of the KNN method, while the choice of K could have a considerable impact on performance on more complicated datasets.

# 7. Conclusion

Across all tested K values, the K-Nearest Neighbors algorithm achieved 100% accuracy, demonstrating exceptional performance on the dataset. The outcomes demonstrate that KNN performs exceptionally well on datasets with distinct class separation. According to the study, choosing the best K number is crucial for maintaining a balance between accuracy and complexity, even if a variety of K values can produce good performance, particularly when working with more difficult datasets.