Iris Flower Classification Using K-Nearest Neighbours

Abstract :

This study used the Python sklearn module to assess how well a K-Nearest Neighbours (KNN) model classified iris blossoms. The model's total accuracy was 100%, and its accuracy for each of the three Iris species—setosa, versicolor, and virginica—was flawless. These findings imply that the model is useful for differentiating between these species according to measures of their petals and sepals. However, there is discussion of restrictions related to the size of the dataset and possible generalizability.

Introduction:

For classification problems, the K-Nearest Neighbours (KNN) algorithm is a popular supervised learning method. In this work, we used KNN to categorize iris blooms, a popular dataset with 150 samples from three different varieties. Four characteristics are used to describe each sample: petal length, petal width, sepal length, and sepal length.

Method:

The research was done using the sklearn Iris dataset. Examining feature distributions and any correlations between characteristics were part of the data exploration process. Using a 75% split of the data, a KNN model was trained with 6 neighbours (n_neighbors=6). The test set used to assess the performance of the model was the remaining 25%. For each Iris species as well as the total, calculations were made for accuracy, precision, recall, and F1-score.

Result:

With regard to the test set, the KNN model's overall accuracy was 100%, accurately categorizing all 38 samples. Subsequent investigation showed that every single Iris species (setosa, versicolor, and virginica) had 100% accuracy.

Discussion:

Based on measurements of the petals and sepals, the KNN model performs exceptionally well, indicating that it is useful in differentiating between Iris species. Still, there are a few restrictions to take into account. Firstly, the model's generalizability to other datasets with various features or more complicated classification issues may be limited by the modest and well-known size of the Iris dataset. Second, although hyperparameter adjustment for the number of Neighbours (k) would have improved performance, this was not investigated in the study. Lastly, no comparison

was done with alternative classification algorithms, such Random Forests or Support Vector Machines (SVM), to see whether they would be superior to KNN in this particular scenario.

Conclusion:

This research showed that a KNN model could classify iris blossoms with 100% accuracy. Before firmly implementing the model to more difficult classification tasks, further research and comparison with alternative techniques are necessary due to the limits of dataset size and possible generalizability.

Reference:

https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html
https://realpython.com/knn-python/
https://www.youtube.com/playlist?list=PLeo1K3hjS3uvCeTYTeyfe0-rN5r8zn9rw

Appendix: