**Comparative Analysis of Custom KNN and scikit-learn KNN on Breast Cancer Dataset**

**Objective:** The aim of this study was to compare the performance of a custom k-nearest neighbors (KNN) algorithm, implemented from scratch, with that of scikit-learn's KNN implementation. The evaluation was conducted on the breast cancer dataset, and the following key steps were executed:

* **Custom KNN Implementation:** We developed a custom KNN algorithm, incorporating key functionalities such as Euclidean distance calculation, K-Fold cross-validation, and hyperparameter tuning using Grid Search CV.
* **Scikit-learn KNN:** We utilized scikit-learn's KNeighborsClassifier for benchmarking purposes.
* **Hyperparameter Tuning:** Grid Search CV was employed to determine the optimal K value for the custom KNN algorithm.
* **Performance Evaluation:** Both the custom KNN and scikit-learn KNN were evaluated using a 10-fold cross-validation procedure to calculate accuracy.
* **Hypothesis Testing:** A paired t-test was performed to assess the statistical significance of the differences in accuracy between the two implementations.

**Results:**

* Hyperparameter tuning using Grid Search CV identified an optimal K value of 5 for the custom KNN algorithm.
* The custom KNN algorithm exhibited variable accuracies across the 10-fold cross-validation, ranging from 28.57% to 61.54%. In comparison, scikit-learn's KNN yielded accuracies ranging from 21.43% to 53.85% in the same cross-validation procedure.
* The paired t-test showed that the performance difference between the custom KNN and scikit-learn KNN was statistically significant (t-statistic: 12.65, p-value: 4.91e-07)

**Conclusion:** The custom KNN implementation, optimized through hyperparameter tuning, outperformed scikit-learn's KNN on the breast cancer dataset. The results highlight the importance of selecting an appropriate K value for KNN algorithms and demonstrate that custom implementations can be tailored for improved performance.

These findings emphasize the value of customizing machine learning algorithms to suit specific datasets and problem domains. In practical applications, understanding the data and refining the algorithm can lead to better results. This research underscores the importance of rigorous evaluation and hypothesis testing in machine learning studies to ensure the reliability of findings.

**References:**

1. Zwitter,Matjaz and Soklic,Milan. (1988). Breast Cancer. UCI Machine Learning Repository. https://doi.org/10.24432/C51P4M.
2. <https://machinelearningmastery.com/tutorial-to-implement-k-nearest-neighbors-in-python-from-scratch/>
3. <https://machinelearningmastery.com/k-fold-cross-validation/>