**Programming Assignment 1**

**Report**

**Hemanth Sukumar Vangala**

**1002118951**

**KNN with 10-Fold Cross Validation**

**Introduction**:

K-Nearest Neighbors (KNN) Algorithm:

The K-Nearest Neighbors (KNN) algorithm is a versatile classification and regression technique. It operates by measuring the similarity between data points and making predictions based on the majority class or average value of the k nearest neighbours. The parameter 'k' denotes the number of neighbours considered in the prediction process. KNN is favoured for its simplicity and intuitive nature, making it an accessible choice for many classification tasks.

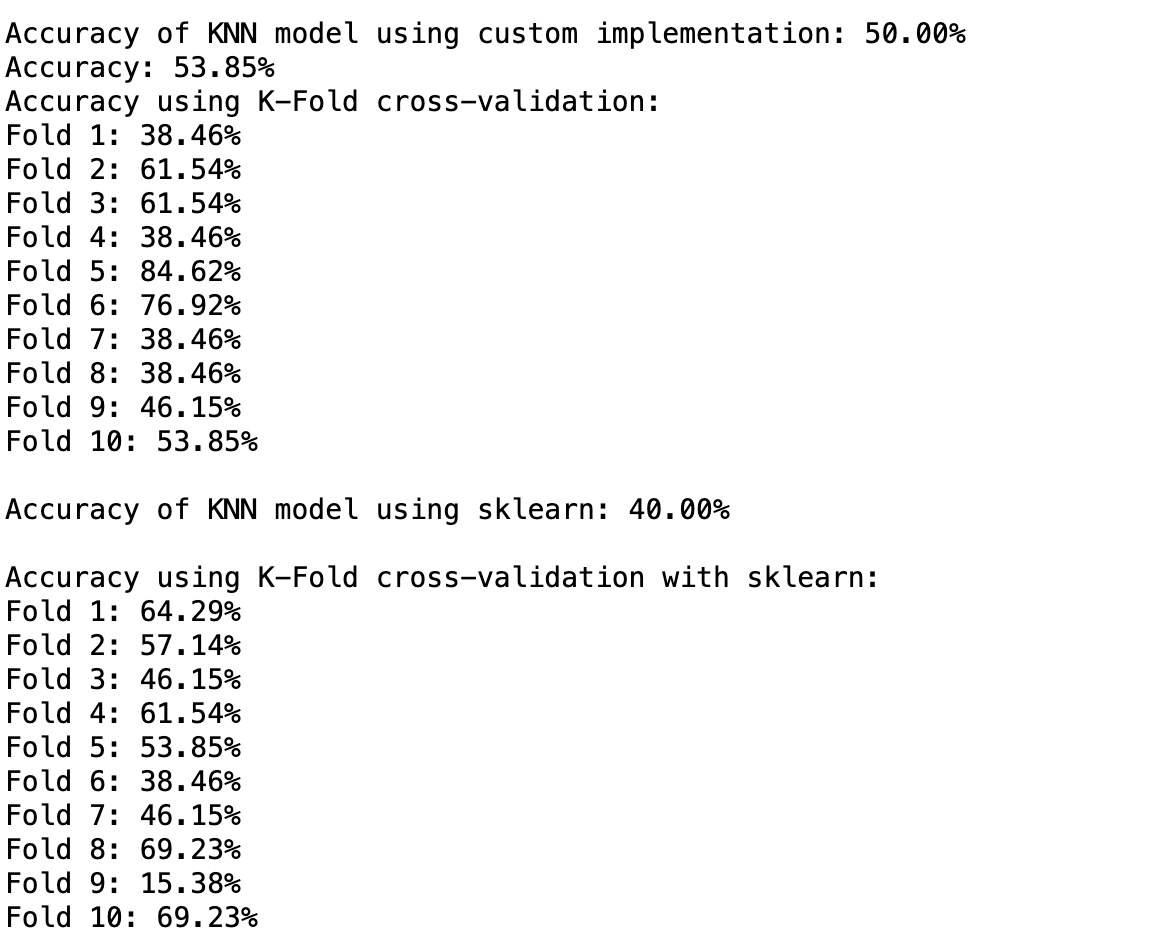
Cross-Validation for KNN:

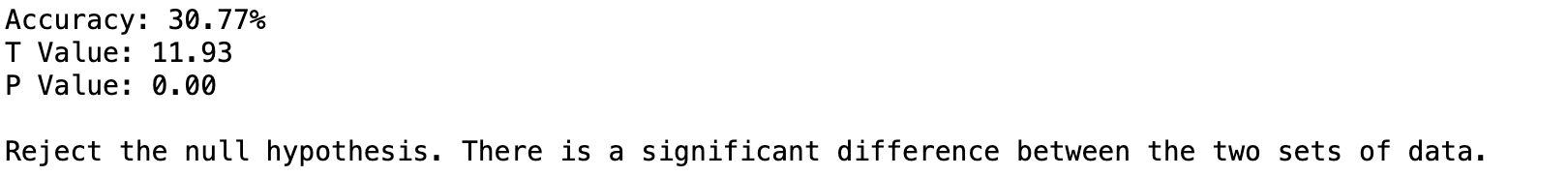
Cross-validation is indispensable to gauge the performance of a KNN classifier and prevent overfitting. In cross-validation, the dataset is divided into k subsets or folds. The algorithm is trained and tested k times, with each fold serving once as the validation set and the remaining folds as the training set. This process, known as k-fold cross-validation, enhances the robustness and reliability of the KNN model.

Performance Comparison and T-Test:

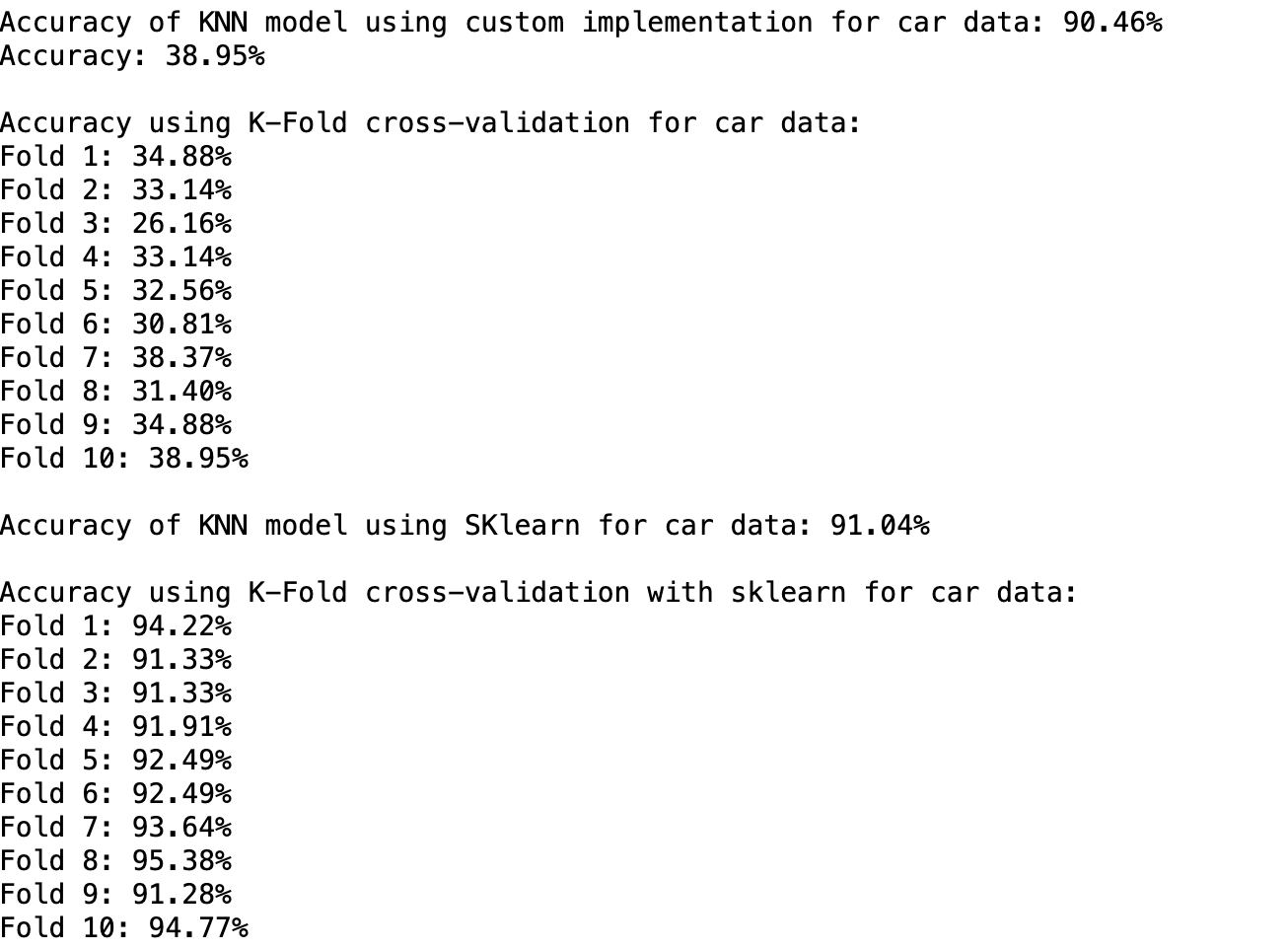
The provided code compares the performance of a custom implementation of the KNN algorithm with the Scikit-learn (Sklearn) implementation. Additionally, it conducts a t-test to determine if there are significant differences in the results between the two approaches. This analysis helps assess both implementations' effectiveness and suitability for the given task.

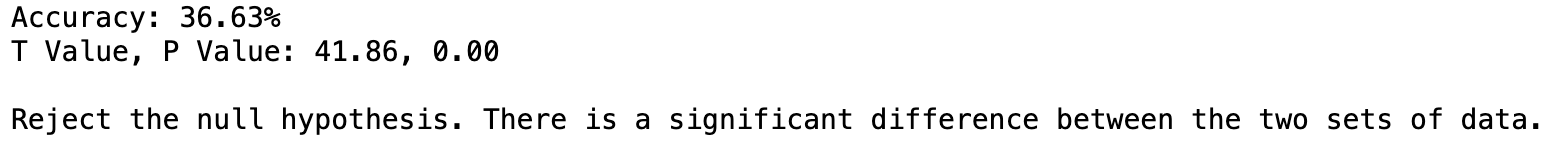
**Hayes Roth dataset output:**





**Car dataset output:**





**Breast cancer dataset output:**

