PROJECT REPORT

ADVANCE MACHINE LEARNING

BY :- NIKHIL REDDY ADDULA

Executive Summary:

The goal of this project was to demonstrate the use of K-Nearest Neighbors (KNN) algorithm in the classification of wine samples. The dataset used in this project was the Wine dataset, which contains 178 samples of wine, each represented by 13 features.

The first step was to load the dataset using the load_wine() function from sklearn.datasets module. The data was then split into training and testing sets using train_test_split() function from sklearn.model_selection module.

A KNN classifier was then created using the KNeighborsClassifier class from the sklearn.neighbors module, and the classifier was fit to the training data. The default parameters of the classifier were used, so no specific parameter was set. Predictions were then made on the training data, and the accuracy score was calculated using the accuracy_score() function from the sklearn.metrics module. The accuracy score for the training data was 0.7958.

The KNN classifier was then re-done with some specific parameters, such as algorithm, leaf_size, metric, p, n_jobs, n_neighbors and weights, and it was fit to the training data. Predictions were then made on the testing data, and the accuracy score was calculated. The accuracy score for the testing data was 0.7222.

Overall, the KNN algorithm was shown to be effective in classifying the wine samples in the Wine dataset, with an accuracy score of 0.7958 on the training data and 0.7222 on the testing data. However, it should be noted that the accuracy score on the testing data is slightly lower than that on the training data, which suggests that the model may be overfitting to the training data. Therefore, further optimization and tuning of the model may be needed to improve its performance on the testing data.