Project Description

Project Overview:

In this project, we will be working with a self-selected dataset to apply the K-Nearest Neighbors (KNN), K-Means, and Naive Bayes algorithms for classification. The objective is to split the dataset into training and testing sets, implement the KNN, K-Means, and Naive Bayes models, and evaluate their performances by printing accuracy and other relevant parameters.

Project Steps:

➢ **Dataset Selection:**

Select a suitable dataset for the project. It can be obtained from various sources, such as public repositories or domain-specific data providers. Ensure that the dataset contains both input features and corresponding labels for classification purposes.

➢ **Data Split:**

Preprocess the dataset by splitting it into three subsets: training set, validation set, and testing set. The training set will be used to train the KNN, K-Means, and Naive Bayes models, while the validation set will be used for fine-tuning and hyperparameter optimization. The testing set will be used to evaluate the final performance.

➢ **KNN Model Implementation:**

Implement the KNN algorithm using appropriate libraries or build it from scratch, depending on your preference and expertise. Adjust the key hyperparameters, such as the value of K (the number of nearest neighbors to consider), to optimize the model's performance.

➢ **K-Means Model Implementation:**

Implement the K-Means algorithm for clustering using appropriate libraries or build it from scratch. This step is optional for classification tasks, but it can provide insights into the structure of the data and aid in feature engineering.

➢ **Naive Bayes Model Implementation:**

Implement the Naive Bayes algorithm using appropriate libraries or build it from scratch. Consider different types of Naive Bayes models (e.g., Gaussian, Multinomial) based on the nature of your dataset.

➢ **Model Training and Testing:**

Train the KNN, K-Means, and Naive Bayes models using the training set obtained in Step 2. Apply the trained models to predict the labels for the validation and testing sets. Compare the predicted labels with the actual labels from the testing set.

➢ **Evaluation and Parameter Printing:**

For all models, calculate the accuracy by comparing the predicted labels with the actual labels from the testing set. Compute other relevant performance metrics, such as precision, recall, and F1-score, if applicable. Print the accuracy and other calculated parameters to provide a comprehensive evaluation of the models' performances.

➢ **Iteration and Improvement:**

If desired, iterate on the project by adjusting different aspects, such as the dataset, preprocessing techniques, hyperparameters, or exploring alternative classification models. Document and analyze the results obtained from different iterations, and make appropriate improvements to enhance the models' performances.

By completing this project, you will gain practical experience in implementing KNN, K-Means, and Naive Bayes algorithms for classification tasks. Additionally, you will learn how to evaluate the performance of the models by measuring accuracy and other relevant parameters, providing valuable insights for further improvements in the field of AI.

Good Luck.