6Assessment: KNN Model Implementation and Evaluation

Assessment Type: Individual, Performance-Based

Desired Skills Assessed:

- Understanding of the KNN algorithm
- Ability to implement the KNN algorithm using a programming language (e.g., Python)
- Ability to evaluate the performance of a KNN model
- Skill in interpreting model results
- Critical thinking and problem-solving skills

Time Allotment: 2 hours

Instructions:

You are tasked with building a K-Nearest Neighbors (KNN) model to classify a given dataset.

Dataset: A dataset containing information about various types of flowers, including features like sepal length, sepal width, petal length, and petal width. The target variable is the species of the flower (Iris-setosa, Iris-versicolor, or Iris-virginica).

Task:

1. **Data Exploration:**
    - Load the dataset into your chosen programming environment (e.g., Python).
    - Explore the data using descriptive statistics and visualizations.
    - Identify any missing values or outliers.
    - Preprocess the data as necessary (e.g., handling missing values, normalization).
2. **Model Building:**
    - Split the dataset into training and testing sets.
    - Implement the KNN algorithm using a suitable library (e.g., scikit-learn).
    - Train the KNN model on the training set.
3. **Model Evaluation:**
    - Evaluate the performance of the model on the testing set using appropriate metrics (e.g., accuracy, confusion matrix).

- Interpret the results and discuss the model's strengths and weaknesses.

4. **Model Improvement (Optional):**
    - Explore techniques to improve the model's performance, such as feature engineering or hyperparameter tuning.
    - Implement these techniques and re-evaluate the model.