K-Nearest Neighbors Classification
BMC PGP AI/ML C3 - Classification

In this assignment, you will implement the K-Nearest Neighbors (KNN) classification algorithm from scratch and apply it to a real-world dataset. The objective is to gain a deep understanding of how KNN works and its practical applications.

Tasks:
1. Implement the KNN algorithm:
    a. Create a Python class for KNN that includes methods for fitting the model, making predictions, and calculating accuracy
        i. Hint: Your class functions could be __init__(self,k), fit(self,x,y), predict(self,x), classification_score(self,y_true,y_pred)
        ii. The distances calculation, sorting to get top K neighbours and voting to get class value can be easily implemented with numpy functions on arrays - that is the direction you should be thinking in
    b. Allow the class to take the value of 'k' as a parameter
2. Working with the dataset:
    a. Load the dataset given with the problem statement
    b. Preprocess the dataset by normalizing the features or performing any necessary data cleaning
    c. Split the dataset into training and testing sets to evaluate the model
    d. Train and evaluate the KNN model and report accuracy on test data
3. Experiment with different 'k' values:
    a. Perform multiple runs of the KNN algorithm with different values of 'k' and observe how it affects the accuracy.
    b. Plot a graph with 'k' on x-axis and classification accuracy on the y-axis

Deliverables:
1. A Jupyter Notebook with the following sections-
    a. A KNN class [code written in simple functions format without constructing it as a Python class will be penalized]
    b. Data split and basic preprocessing for the algorithm
    c. Initial training and accuracy report for one chosen k-value
    d. Experiments with different k values and comparison graph. Report some analysis here on your insights from this experiment
2. Grading: Students will be assessed based on the correctness and efficiency of their KNN implementation, their ability to apply it to a real dataset, and the quality of their analysis and experimentation with different 'k' values. This assignment will help students

understand the KNN algorithm, data preprocessing, model evaluation, and the importance of parameter tuning in machine learning.