ACADEMIC CITY UNIVERSITY COLLEGE

(Affiliated to the University of Mines and Technology, Takwa, Ghana)

Artificial Intelligence

Machine Learning

Date: 25/04/2023
Submission: 6:30 pm

Data Analysis

a. Load the iris dataset into a Pandas data frame and display the first 5 rows.

b. Display the shape of the dataset (number of rows and columns).

c. Display the summary statistics (mean, standard deviation, minimum, and maximum) for each column.

d. Create a new column in the data frame that is the product of the Sepal Length and Sepal Width columns.

e. Display the unique values in the species column.

f. Create a scatterplot of the sepal length vs. sepal width columns, with different colours for each species.

Machine Learning (Classification)

a. Split the dataset into training and testing sets (70% training, 30% testing).

b. Implement a logistic regression model to predict the species based on the sepal length and sepal width columns. Evaluate the model using accuracy, precision, recall, and the F1 score.

c. Implement a k-nearest neighbours classifier with k = 3 to predict the species based on all four columns. Evaluate the model using accuracy, precision, recall, and the F1 score. d. Implement a

decision tree classifier to predict the species based on all four columns. Evaluate the model using accuracy, precision, recall, and the F1-score.

e. Implement a support vector machine (SVM) classifier to predict the species based on all four columns. Evaluate the model using accuracy, precision, recall, and the F1 score.

Machine Learning (Clustering)

a. Implement a k-means clustering algorithm to cluster the data based on the sepal length and sepal width columns. Visualize the clusters with different colors.

b. Implement a hierarchical clustering algorithm to cluster the data based on all four columns. Visualize the clusters with different colors.

c. Implement a DBSCAN clustering algorithm to cluster the data based on all four columns. Visualize the clusters with different colors.

d. Implement a Gaussian Mixture Model clustering algorithm to cluster the data based on all four columns. Visualize the clusters with different colors.

e. Evaluate the performance of each clustering algorithm using the silhouette score.

Good luck!