Assignment Code: DA-AG-009

Supervised Classification: Decision Trees, SVM, and Naive Bayes| Assignment

Instructions: Carefully read each question. Use Google Docs, Microsoft Word, or a similar tool to create a document where you type out each question along with its answer. Save the document as a PDF, and then upload it to the LMS. Please do not zip or archive the files before uploading them. Each question carries 20 marks.

Total Marks: 200

Question 1 : What is Information Gain, and how is it used in Decision Trees?

Answer:

Question 2: What is the difference between Gini Impurity and Entropy?

Hint: Directly compares the two main impurity measures, highlighting strengths, weaknesses, and appropriate use cases.

Answer:

Question 3:What is Pre-Pruning in Decision Trees?

Answer:

Question 4:Write a Python program to train a Decision Tree Classifier using Gini Impurity as the criterion and print the feature importances (practical).

Hint: Use criterion='gini' in DecisionTreeClassifier and access .feature_importances_.

(Include your Python code and output in the code box below.)

Answer:

Question 5: What is a Support Vector Machine (SVM)?

Answer:

Question 6: What is the Kernel Trick in SVM?

Answer:

Question 7: Write a Python program to train two SVM classifiers with Linear and RBF kernels on the Wine dataset, then compare their accuracies.

Hint:Use SVC(kernel='linear') and SVC(kernel='rbf'), then compare accuracy scores after fitting on the same dataset.

(Include your Python code and output in the code box below.)

Answer:

Question 8: What is the Naïve Bayes classifier, and why is it called "Naïve"?

Answer:

```




```

Question 9: Explain the differences between Gaussian Naïve Bayes, Multinomial Naïve Bayes, and Bernoulli Naïve Bayes

Answer:

```




```

Question 10: **Breast Cancer Dataset**

Write a Python program to train a Gaussian Naïve Bayes classifier on the Breast Cancer dataset and evaluate accuracy.
Hint:Use GaussianNB() from sklearn.naive_bayes and the Breast Cancer dataset from sklearn.datasets.
(Include your Python code and output in the code box below.)

Answer:

```




```