Machine Learning Laboratory – Assignment IV

Data Set

This database contains 14 attributes. The "Diseased" (target variable) field indicates the presence of disease in the patient. It is an integer value where 0 represents no presence, and 1 represents the presence of the disease.

A) Decision Tree Model

Build a Decision Tree model using `scikit-learn` to classify patients with and without disease based on the provided dataset.

1. Deal with the missing data and correctly encode the categorical variables in the dataset.

2. Build a classification tree using the Gini index as the criterion.

3. Plot the confusion matrix.

4. Find the cross-validation score and plot it.

B) Random Forest Model

Build a Random Forest model to classify patients with and without disease based on the provided dataset.

1. Build a Random Forest model.

2. Print the feature importance array.

3. Plot the confusion matrix.

4. Find the cross-validation score and plot it.

5. Print the classification score.

6. Find the optimal parameters for the model:

   (a) Optimal number of trees.
   (b) Optimal maximum depth.
   (c) Optimal number of features.