BIN2023R01 – INTRODUCTION TO DATAMINING & MACHINE LEARNING FOR BIOINFORMATICS

Lab Exercise 7- Classification with Decision Tree

Aim: To build a decision tree model that can classify and predict the given dataset.

Procedure:

1. Import the required packages for model construction. (https://scikit-learn.org/stable/modules/tree.html)
2. Conduct data preprocessing and select features.
3. Build the decision tree model.
4. Evaluate the performance of the model.
5. Make predictions for user-defined data.
6. Contrast the performance metrics of the decision tree model with the logistic regression model.

Questions:

1. What is a decision tree model, and how does it work? What are the advantages of using decision tree models for classification tasks?
2. How does a decision tree split the data to make predictions? What are the key hyperparameters in a decision tree model, and how do they affect the model's performance?
3. Can decision tree models handle categorical and numerical features differently? If so, how?
4. What is pruning in decision tree models, and why is it important?
5. Can decision tree models handle multi-class classification tasks, or are they limited to binary classification?
6. What is the definition of the Gini index?
7. Interpret which model exhibits higher accuracy and lower error rates relative to the other.
8. If the decision tree model exhibits low accuracy, what strategies could be employed to enhance its performance?

Soft copy deadline: March 11th 11:59PM
Hard copy deadline: March 12th 3:15PM