

Data Science, A1

Due Date: 4th November, 2024

For the dataset drug200, build a classification model using the Decision Tree, and report your findings, with some discussion on the results. You should change the parameters of the DT and split ratio of Training and Testing Datasets. In all it should report 5 different result sets.

About Dataset:

Imagine that you are a medical researcher compiling data for a study. You have collected data about a set of patients, all of whom suffered from the same illness. During their course of treatment, each patient responded to one of 5 medications, Drug A, Drug B, Drug c, Drug x and y.

Your job is to build a model to find out which drug might be appropriate for a future patient with the same illness. The features of this dataset are Age, Sex, Blood Pressure, and the Cholesterol of the patients, and the target is the drug that each patient responded to.

It is a sample of multiclass classifier, and you can use the training part of the dataset to build a decision tree, and then use it to predict the class of a unknown patient, or to prescribe a drug to a new patient.