CA 5314: Practice Exercise 3

Decision Tree

Aim:

It shows how to build and optimize Decision Tree Classifier of "Diabetes dataset" using Python Scikit-learn package.

1. Importing Required Libraries Let's first load the required libraries.
2. Loading Data Let's first load the required Diabetes dataset using pandas read CSV function. You can download the data here (datasets_set.csv)
3. Feature Selection Here, you need to divide given columns into two types of variables dependent (or target variable) and independent variable (or feature variables).
4. Splitting Data To understand model performance, dividing the dataset into a training set and a test set is a good strategy.
5. Let's split the dataset by using function train_test_split(). You need to pass 3 parameters features, target, and test_set size.
6. Building Decision Tree Model Let's create a Decision Tree Model using Scikit-learn.
7. Evaluating Model Let's estimate, how accurately the classifier or model can predict the type of cultivars.
8. Accuracy can be computed by comparing actual test set values and predicted values.