Decision Tree on Pima Indians Diabetes Database

Objective:

- To understand the structure of the dataset
- To create a Decision Tree Model

Pima Indians Diabetes Dataset:

The details regarding this dataset are present in the data dictionary

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Lab Environment: Jupyter Notebook

Domain: Medical

Tasks to be performed:

- ✓ Read the .csv file and understand the structure of the dataset.
- ✓ Divide the given columns into two types of variables dependent (or target variable) and independent variable (or feature variables).

Building simple logistic models:

- ✓ To understand model performance, dividing the dataset into a training set and a test set
- ✓ Import the Logistic Regression module and create a Logistic Regression classifier object using Logistic Regression () function.
- ✓ Fit your model on the train set using fit () and perform prediction on the test set using predict ().
- ✓ Evaluate your Model using Confusion Matrix
- ✓ Visualize the confusion matrix using Heatmap.
- ✓ Evaluate the model using model evaluation metrics such as accuracy, precision, and recall.
- ✓ Plot a ROC Curve for your Model