NAME : VARSHINI R
ROLL NO : 18BCS073
DATE : 23/06/2021

Data Warehousing and Data Mining-End Semester Practicals

1. Download a suitable dataset for classification from any Repository. List the attributes and its type in a word Doc.

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

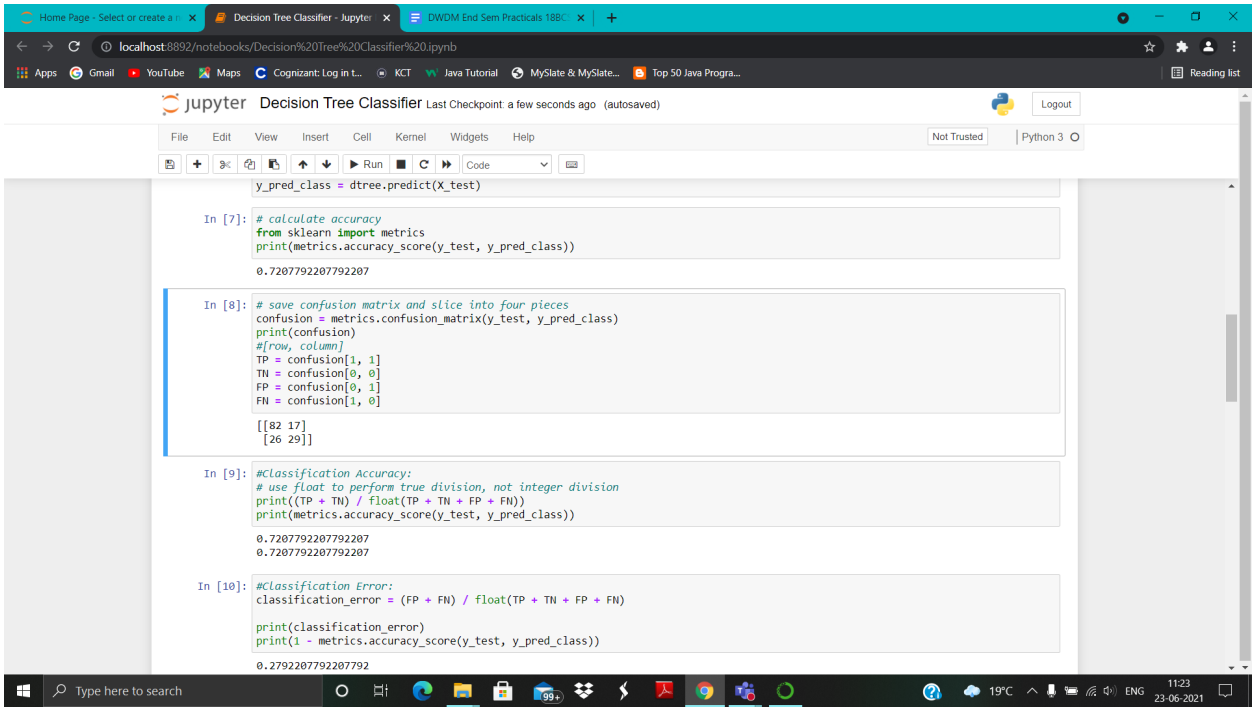
The datasets consist of several medical predictor variables and one target variable, `Outcome.` Predictor variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

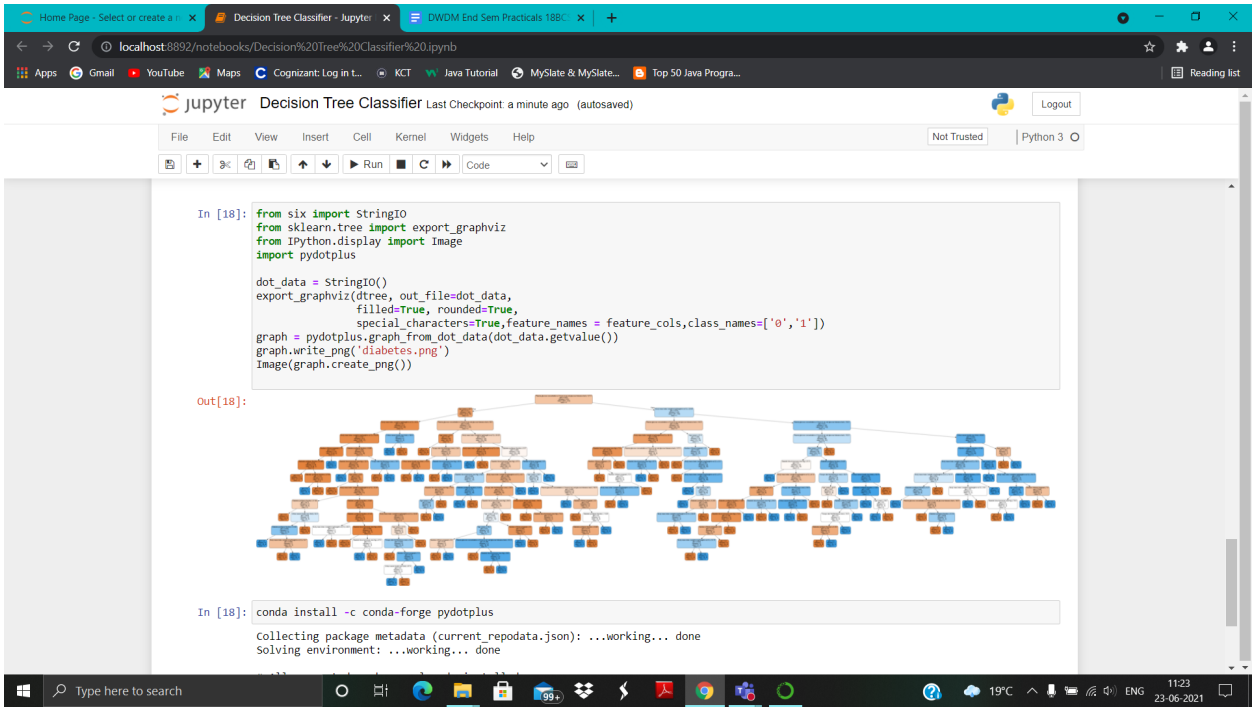
Attribute types	Integer, Real
Instances	768
Attributes	8

Dataset Link -

https://www.kaggle.com/uciml/pima-indians-diabetes-database

2. Load the dataset and set the target and feature variables. Split the dataset into training and test dataset. Build a decision tree classifier with Entropy criteria. Perform Prediction for test dataset using Entropy and print the results in the form of confusion matrix, accuracy and classification report. visualize the decision tree.





3. Upload in your github account. Provide the link for access.

https://github.com/Varshini11122000/End-Semester-Practicals-18BCS073/upload/main