Name: Kritya Shree S
Roll no: 18BCS093
Date: 23/6/2021
Class: III rd CSE - A
Subject: Data Warehousing and Data Mining - U18CSI6203L

End Semester Lab

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

Dataset:

pima-indians-diabetes

Pregnencies	Glucose	BP	Skin_Thickness	Insulin	BMI	Diabetes_Pedigree	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31.0	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0.0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38.0	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30.0	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1
7	107	74	0	0	29.6	0.254	31	1
1	103	30	38	83	43.3	0.183	33	0
1	115	70	30	96	34.6	0.529	32	1
3	126	88	41	235	39.3	0.704	27	0
8	99	84	0	0	35.4	0.388	50	0
7	196	90	0	0	39.8	0.451	41	1
9	119	80	35	0	29.0	0.263	29	1
11	143	94	33	146	36.6	0.254	51	1

Dataset Description:

- **Dataset Used:** Pima-indians-diabetes.csv

This dataset can be used to predict whether a patient is diabetic or not

- **Attributes present in the dataset:**

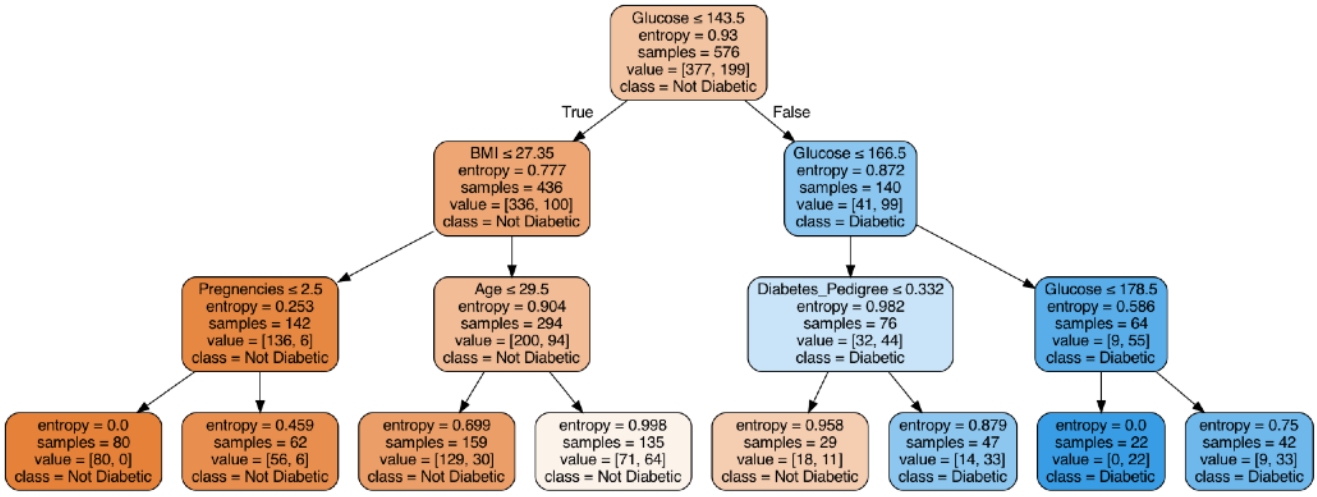
Pregnancies: Number of times the patient was pregnant
Glucose: Plasma Glucose Concentration
BP: Diastolic blood pressure measured in mm Hg
Skin_Thickness: Triceps skin fold thickness measured in mm
Insulin: Serum Insulin measured in muU/ml
BMI: Body Mass Index of the patient
Diabetes_Pedigree: Likelihood score of diabetes based on family history
Age: Age of the patient
Outcome: If the patient is diabetic or not (1- Diabetic, 0 - Non-Diabetic)

Attribute Description:

Pregnancies: Numeric Data (Discrete)
Glucose: Numeric Data (Continuous)
BP: Numeric Data (Continuous)
Skin_Thickness: Numeric Data (Discrete)
Insulin: Numeric Data (Continuous)
BMI: Numeric Data (Continuous)
Diabetes_Pedigree: Numeric Data (Continuous)
Age: Numeric Data (Discrete)
Outcome: Categorical Data (Ordinal)

2. Load the dataset and set the target and feature variables. Split the dataset into training and test dataset. Build 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.

Decision Tree Visualisation:



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

https://github.com/KrityaShree/DWDM-End-Semester-Lab