

Name: Lashya. P

Roll no: 18BCS045

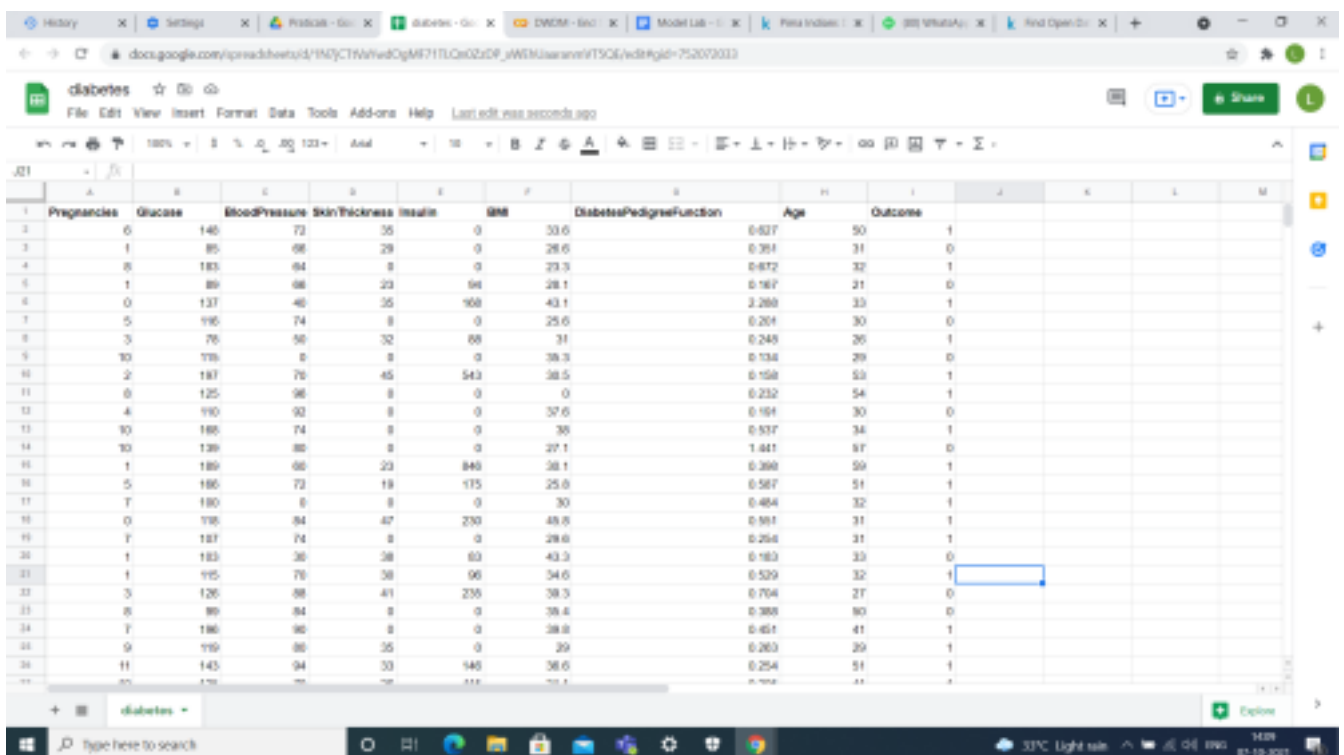
Date: 07-10-2021

Subject: PROTO21 Data Warehousing and Data Mining - U18CSI6203L

Model Exam

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

Dataset:



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome				
2	0	140	72	35	0	33.6	0.627	50	1				
3	1	85	68	29	0	26.6	0.351	31	0				
4	0	183	64	0	0	23.3	0.672	32	1				
5	1	89	66	23	94	28.1	0.167	21	0				
6	0	137	40	35	168	43.1	2.268	33	1				
7	5	190	74	0	0	25.6	0.201	30	0				
8	3	78	58	32	66	31	0.245	26	1				
9	10	178	0	0	0	38.3	0.136	29	0				
10	2	187	70	45	543	38.5	0.158	53	1				
11	0	125	96	0	0	0	0.232	54	1				
12	4	190	92	0	0	37.6	0.191	30	0				
13	10	168	74	0	0	38	0.537	34	1				
14	10	138	80	0	0	27.1	1.441	31	0				
15	1	189	66	23	846	38.1	0.369	59	1				
16	5	166	72	19	175	25.0	0.567	51	1				
17	7	180	0	0	0	30	0.484	32	1				
18	0	178	64	47	230	48.5	0.591	31	1				
19	7	187	74	0	0	28.6	0.264	31	1				
20	1	183	36	38	63	43.3	0.183	33	0				
21	1	195	70	38	96	34.6	0.529	32	1				
22	3	126	68	41	235	38.3	0.704	27	0				
23	8	89	64	0	0	38.4	0.388	30	0				
24	7	186	90	0	0	38.0	0.451	41	1				
25	9	179	86	35	0	29	0.263	29	1				
26	11	143	94	33	146	38.6	0.254	51	1				
27	0	178	70	38	112	33.1	0.178	31	0				

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 mmHg

Skin_Thickness: Triceps skinfold 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)

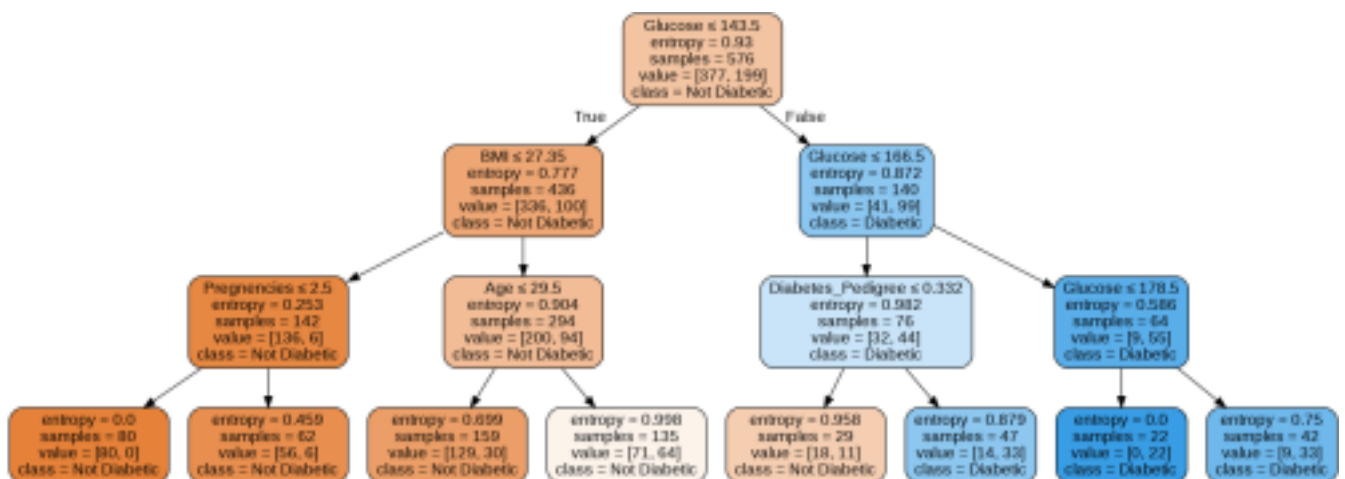
Diabetes_Pedigree: Numeric Data (Continuous)

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 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.

Decision Tree Visualisation:



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

<https://github.com/Lashya13/DWDM-Model-Exam>