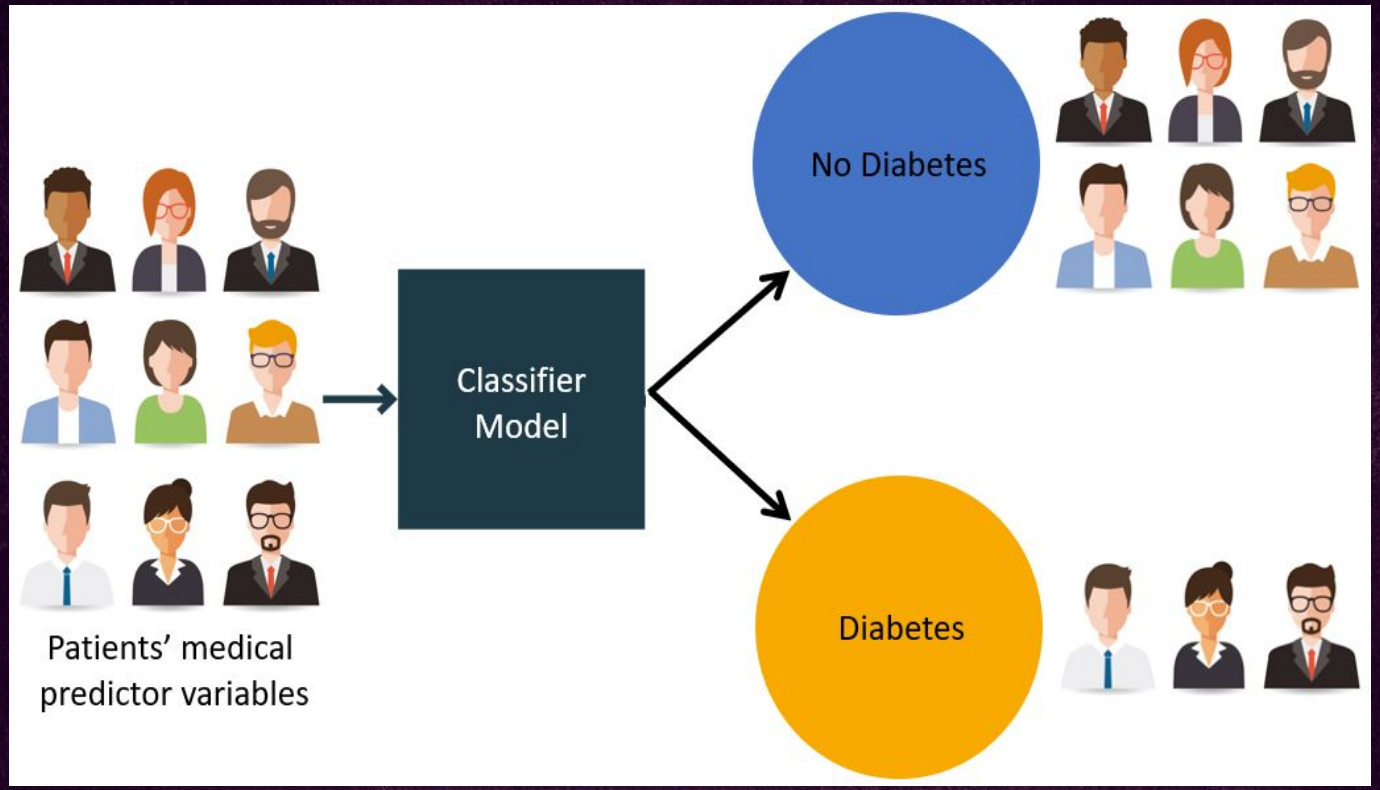
DIABETES PREDICTION



OBJECTIVE

TO BUILD A PREDICTIVE MODEL WHICH CLASSIFIES WHETHER A PERSON IS DIABETIC OR NOT BASED ON THE PARAMETERS

1. PREGNANCIES
2. GLUCOSE
3. BLOOD PRESSURE
4. SKIN THICKNESS
5. INSULIN
6. BMI
7. DIABETES PEDIGREE FUNCTION
8. AGE

DIABETES.CSV

	A	B	C	D	E	F	G	H	I
1	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
2	6	148	72	35	0	33.6	0.627	50	1
3	1	85	66	29	0	26.6	0.351	31	0
4	8	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	2.288	33	1
7	5	116	74	0	0	25.6	0.201	30	0
8	3	78	50	32	88	31	0.248	26	1
9	10	115	0	0	0	35.3	0.134	29	0
10	2	197	70	45	543	30.5	0.158	53	1
11	8	125	96	0	0	0	0.232	54	1
12	4	110	92	0	0	37.6	0.191	30	0
13	10	168	74	0	0	38	0.537	34	1
14	10	139	80	0	0	27.1	1.441	57	0
15	1	189	60	23	846	30.1	0.398	59	1
16	5	166	72	19	175	25.8	0.587	51	1
17	7	100	0	0	0	30	0.484	32	1
18	0	118	84	47	230	45.8	0.551	31	1
19	7	107	74	0	0	29.6	0.254	31	1
20	1	103	30	38	83	43.3	0.183	33	0
21	1	115	70	30	96	34.6	0.529	32	1
22	3	126	88	41	235	39.3	0.704	27	0
23	8	99	84	0	0	35.4	0.388	50	0
24	7	196	90	0	0	39.8	0.451	41	1
25	9	119	80	35	0	29	0.263	29	1
26	11	143	94	33	146	36.6	0.254	51	1
27	10	125	70	26	115	31.1	0.205	41	1
28	7	147	76	0	0	39.4	0.257	43	1
29	1	97	66	15	140	23.2	0.487	22	0

STEPS INVOLVED

1. IMPORTING DATASET
2. ANALYSE THE DATA SET
3. SPLITTING UP OF DATA
4. APPLYING ML ALGORITHM
5. EVALUATION OF MODEL

IMPORTING DATASET

THE BASIC PROCESS OF LOADING DATA FROM A CSV FILE INTO A PANDAS DATAFRAME (WITH ALL GOING WELL) IS ACHIEVED USING THE "READ_CSV" FUNCTION IN PANDAS

```
IMPORT PANDAS AS PD
DATA = PD.READ_CSV("FILENAME.CSV")
```

ANALYSING DATASET

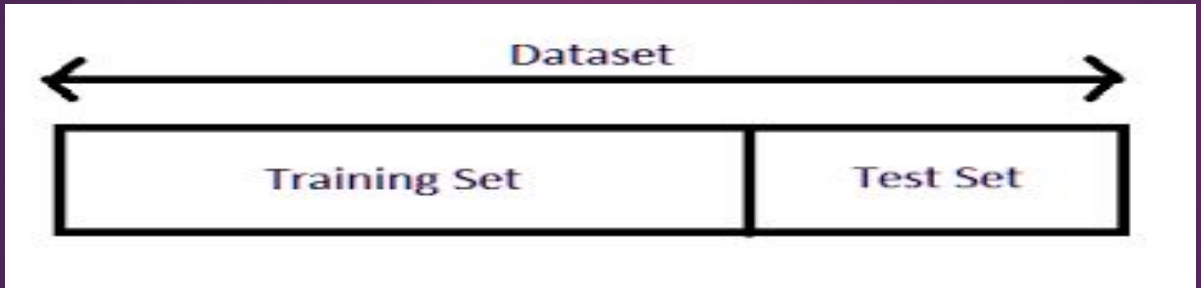
Basically we need to analyse and remove the columns which are not necessary and use the ones which are necessary and handle the missing values if any

Splitting up of data

REQUIREMENTS:

- IMPORT TRAIN_TEST_SPLIT FROM SKLEARN.MODEL_SELECTION

DATA IS SPLITTED INTO TRAINING AND TESTING SET BASED ON SIZE OF THE DATASET .



Applying ML algorithm

CLASSIFICATION IS THE PROCESS OF PREDICTING THE CLASS OF GIVEN DATA POINTS. CLASSIFICATION PREDICTIVE MODELING IS THE TASK OF APPROXIMATING A MAPPING FUNCTION (F) FROM INPUT VARIABLES (X) TO DISCRETE OUTPUT VARIABLES (Y).

1. LINEAR CLASSIFIERS: LOGISTIC REGRESSION
2. NEAREST NEIGHBOR
3. RANDOM FOREST

EVALUATION OF MODEL

Accuracy is a common evaluation metric for classification problems.
It's the number of correct predictions made as a ratio of all predictions made

REQUIREMENTS:

- import accuracy_score from sklearn.metrics