

ML-MINOR-MAY PROJECT DESCRIPTION

1. Problem Statement:-

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.

2. Aim Of the Project:-

To build a machine learning model to accurately predict whether or not the patients in the dataset have diabetes or not.

3. List of Libraries used in my project:-

- I. numpy
- II. pandas
- III. matplotlib.pyplot
- IV. seaborn
- V. os
- VI. scipy

The platform I have used for my machine learning project is Google Colaboratory and have used list of libraries which are mentioned above and I have also imported sklearn for the creating my method and have used Logistic Regression (Algorithm).

4. Explanation:-

From the given dataset our aim is to predict the number of patients who have diabetes or not and list of patients who are having diabetes are Indicated by 1 and patients who are not having diabetes are Indicated by

0. Our model shows how many patients are affected by diabetes and how many are not affected and they are indicated by histogram . each patient have been shown there glucose level, Bloodpressure, Insulin, BMI, skinthickness

- In the first quadrant Our model is indicating 98 as TruePositive which means 98 peoples are having diabetes and our model is showing a positive result for all 98 peoples.
- In the second our model indicating 9 peoples as FalsePositive which is slight dangerous.
- In the third 18 peoples are shown as TrueNegative which is more dangerous than FalsePositive.actually 18 peoples are having diabetes but our model showing they don't have diabetes.
- Finally in the fourth 29 peoples are shown as FalseNegative where our model predicts the correct value and indicated 29 peoples are not having diabetes

Finally our accuracy score is 82.467 which means our models holds good and indicating the correct number of patients who are having diabetes and are not having diabetes.

5. Conclusion:-

Finally we are getting a accuracy score of 82.46% using Logistic Regression. Hence I have to conclude that our model predicted the correct values

Classification Report is:

precision	recall	f1-score	support
-----------	--------	----------	---------

0	0.84	0.92	0.88	107
1	0.76	0.62	0.68	47

accuracy			0.82	154	
macro avg	0.80	0.77	0.78	154	
weighted avg		0.82	0.82	0.82	154