

WRITE_UP FOR CAPASTONE PROJECT: -

Healthcare Project: - 2

DESCRIPTION Of DATA - SET

The datasets consists of several medical predictor variables and one target variable (Outcome). Predictor variables includes the number of pregnancies the patient has had, their BMI, insulin level, age, and more.

Variables	Description
Pregnancies	Number of times pregnant
Glucose	Plasma glucose concentration in an oral glucose tolerance test
Blood Pressure	Diastolic blood pressure (mm Hg)
Skin Thickness	Triceps skinfold thickness (mm)
Insulin	Two hour serum insulin
BMI	Body Mass Index
Diabetes Pedigree Function	Diabetes pedigree function
Age	Age in years
Outcome	Class variable (either 0 or 1). 268 of 768 values are 1, and the others are 0

Project Task Performed

Data Exploration:

1. Check missing value in whole dataset as well in following columns and found there is no missing value but in description written value having zero treated as missing value.

- Glucose
- Blood Pressure
- Skin Thickness
- Insulin
- BMI

2. Drawn Histogram for all variables.

Data Exploration:

1. Check balancing of data and found data is imbalance try to balance data.
2. Scatter plot made for two relative variables for understand the relationship.
3. Performed correlation analysis, & drawn Heat map.

Data Modeling:

1. Apply Seven classification algorithm to build model. & Compare all models with the results from KNN algorithm.
2. Created a classification report by analyzing AUC (ROC curve), Classification reports describe to explain what values of these parameter i have used.

Tableau Work done :-

Data Reporting:

2. Created a dashboard in tableau by choosing appropriate chart types and metrics useful for the business.

The dashboard have the following:

- a. Pie chart to describe the diabetic or non-diabetic population.
- b. Scatter charts between relevant variables to analyze the relationships.
- c. Histogram charts to analyze the distribution of the data.
- d. Heatmap of correlation analysis among the relevant variables.
- e. Created bins of these age values: 20-25, 25-30, 30-35, etc.
- f. Created bubble chart to Analyze different variables.

THANK YOU

Project Done By
Aditya Mani