Context

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.

Content

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 so on.

Acknowledgments

Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., & Johannes, R.S. (1988). [Using the ADAP learning algorithm to forecast the onset of diabetes mellitus](http://rexa.info/paper/04587c10a7c92baa01948f71f2513d5928fe8e81). *In Proceedings of the Symposium on Computer Applications and Medical Care* (pp. 261--265). IEEE Computer Society Press.

About this file

The datasets consist of several medical predictor (independent) variables and one target (dependent) variable, Outcome. Independent variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

Columns

Pregnancies

Number of times pregnant

Glucose

Plasma glucose concentration a 2 hours in an oral glucose tolerance test

BloodPressure

Diastolic blood pressure (mm Hg)

SkinThickness

Triceps skin fold thickness (mm)

Insulin

2-Hour serum insulin (mu U/ml)

BMI

Body mass index (weight in kg/(height in m)^2)

DiabetesPedigreeFunction

Diabetes pedigree function

Age

Age (years)

Outcome

Class variable (0 or 1) 268 of 768 are 1, the others are 0