Dataset: DiabetesDiagnosis

Attributes: 8 numeric
Class: Nominal (binary)
Instances: 768
Missing Values: None

The binary-valued diagnostic variable presented in this dataset relates whether a patient shows signs of diabetes according to World Health Organization criteria (i.e., if the 2 hour post-load plasma glucose was at least 200 mg/dl at any survey examination or if found during routine medical care).   The population lives near Phoenix, Arizona, USA.

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.

Each instance represents individual patients and their various medical attributes along with diabetes classification

Attributes)all nominal):
1. Pregnancies: Number of pregnancies
2. PG Concentration: Plasma glucose at 2 hours in an oral glucose tolerance test
3. Diastolic BP: Diastolic Blood Pressure (mm Hg)
4. Tri Fold Thick: Triceps Skin Fold Thickness (mm)
5. Serum Ins: 2-Hour Serum Insulin (mu U/ml)
6. BMI: Body Mass Index: (weight in kg/ (height in m)^2)
7. DP Function: Diabetes Pedigree Function
8. Age: Age (years)

Class: Diagnosis = sick, healthy

Reference:
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. Proceedings of the Symposium on Computer Applications and Medical Care (pp. 261—265), IEEE Computer Society Press.