

The “australian” dataset hosted on (UCI’s Machine Learning Repository) contains approximately 690 observations, with 15 variables. The dependent variable that in all cases we will be trying to predict is whether or not an “individual” could get approved in the credit application.

This file concerns credit card applications. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. This dataset is interesting because there is a good mix of attributes -- continuous, nominal with small numbers of values, and nominal with larger numbers of values. There are also a few missing values.

There are 6 numerical and 8 categorical attributes. The original data label has been changed for the convenience of the statistical algorithms. For example, attribute 4 originally had 3 labels: p, g, gg. And these have been changed to labels 1, 2, 3.

Attributes(15):

A1: Categorical. Values: 0, 1.

A2: Continuous. Range from: 13.75-80.25.

A3: Continuous. Range from: 0.0-28.0.

A4: Categorical. Values: 1, 2, 3.

A5: Categorical. Values: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14.

A6: Categorical. Values: 1, 2, 3, 4, 5, 6, 7, 8, 9.

A7: Continuous. Range from: 0.0-28.5.

A8: Categorical. Values: 0, 1.

A9: Categorical. Values: 0, 1.

A10: Continuous. Range from: 0-67.

A11: Categorical. Values: 0, 1.

A12: Categorical. Values: 1, 2, 3.

A13: Continuous. Range from: 0-2000.

A14: Continuous. Range from: 1-100001.

A15: Binary. Values: 1, 2.