Description of the Dataset:

THIS CREDIT DATA ORIGINATES FROM QUINLAN (see below).

1. Title: Australian Credit Approval

2. Sources:

(confidential)

Submitted by quinlan@cs.su.oz.au

3. Past Usage:

See Quinlan,

\* "Simplifying decision trees", Int J Man-Machine Studies 27,

Dec 1987, pp. 221-234.

\* "C4.5: Programs for Machine Learning", Morgan Kaufmann, Oct 1992

4. Relevant Information:

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.

5. Number of Instances: 690

6. Number of Attributes: 14 + class attribute

7. Attribute Information: THERE ARE 6 NUMERICAL AND 8 CATEGORICAL ATTRIBUTES.

THE LABELS HAVE 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.

A1: 0,1 CATEGORICAL

a,b

A2: continuous.

A3: continuous.

A4: 1,2,3 CATEGORICAL

p,g,gg

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

ff,d,i,k,j,aa,m,c,w, e, q, r,cc, x

A6: 1, 2,3, 4,5,6,7,8,9 CATEGORICAL

ff,dd,j,bb,v,n,o,h,z

A7: continuous.

A8: 1, 0 CATEGORICAL

t, f.

A9: 1, 0 CATEGORICAL

t, f.

A10: continuous.

A11: 1, 0 CATEGORICAL

t, f.

A12: 1, 2, 3 CATEGORICAL

s, g, p

A13: continuous.

A14: continuous.

A15: 1,2

+,- (class attribute)

8. Missing Attribute Values:

37 cases (5%) HAD one or more missing values. The missing

values from particular attributes WERE:

A1: 12

A2: 12

A4: 6

A5: 6

A6: 9

A7: 9

A14: 13

THESE WERE REPLACED BY THE MODE OF THE ATTRIBUTE (CATEGORICAL)

MEAN OF THE ATTRIBUTE (CONTINUOUS)

9. Class Distribution

+: 307 (44.5%) CLASS 2

-: 383 (55.5%) CLASS 1

10. There is no cost matrix.