1. นายอิทธิพร แก้วอำไพ 590610681
2. Title: Credit Approval

Data types: Multivariate

Instances: 690

Attribute Types: Categorical, Integer, Real

Attributes: 15 + 1(class attribute)

Attribute Information:

A1 – b, a.

A2 – continuous.

A3 – continuous.

A4 – u, y, l, t.

A5 – g, p, gg.

A6 – c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff.

A7 – v, h, bb, j, n, z, dd, ff,o

A8 – continuous.

A9 – t, f.

A10 – t, f.

A11 – continuous.

A12 – t, f.

A13 – g, p, s.

A14 – continuous.

A15 – continuous.

A16 – (+, -) (class attribute)

Missing Attribute Values:

37 cases (5%) have one or more missing values. The missing values from particular attributes are:

A1: 12

A2: 12

A4: 6

A5: 6

A6: 9

A7: 9

A14: 13

Class Distribution:

+: 307 (44.5%)

-: 383 (55.5%)

Source: Submitted by [quinlan@cs.su.oz.au](mailto:quinlan@cs.su.oz.au)

1. ขั้นตอนการทำ pre processing

* จัดการ row ที่มี missing values ตั้งแต่ 20% ขึ้นไปโดยการตัดทั้ง row เนื่องจากมี missing values มากเกินไป โดยจะมีดังนี้ ข้อมูลที่ 207(5), 271(5), 331(5), 457(5), 480(3), 593(5), 602(3), 623(5)
* จัดการเติม missing values ของ A1 โดยการเติม highest frequency value(b) เข้าไปในทุกที่ที่เป็น missing values(A1) เพื่อที่จะได้ไม่ทำให้ข้อมูลเพี้ยนจากเดิม

โดยจะมีดังนี้ ข้อมูลที่ 249, 328, 347, 375, 454, 490, 521, 599, 642, 674

* เติม missing values ของ A2 โดยเติมตาม mean โดย mean = 31.5247 เพื่อไม่ให้ค่า mean เปลี่ยนไปจากเดิม โดยเติมข้อมูลที่ 84, 87, 93, 98, 254, 285, 328, 443, 448, 496, 511, 602
* เติม missing values ของ A6 และ A7 ของข้อมูลที่ 535 โดยค่าที่เติมไปคือ c และ v ตามลำดับโดยนำค่ามาจาก mod ของแต่ละ column
* เติม missing values ของ A14 โดนการนำ mean ไปแทนทุกค่าที่หายไป