1.)

ก. gini index = Entropy = 1

ข. 1 – 20

Gini index = Entropy =

Information gain =

ค. M

Gini index = Entropy =

Information gain =

F

Gini index = Entropy =

Information gain =

ง. Family

Gini index = Entropy =

Information gain =

Sport

Gini index = Entropy =

Information gain =

Luxury

Gini index = Entropy =

Information gain =

จ. Small

Gini index = Entropy =

Information gain =

Medium

Gini index = 1 Entropy =

Information gain =

Large

Gini index = Entropy =

Information gain =

Extra large

Gini index = Entropy =

Information gain =

ฉ. Customer ID

ช. เหมาะ เพราะ มีความบริสุทธิ์มากที่สุด

4. Cost(internal) = log m Cost(leaf) = log 3

Tree1 Tree2

Cost(tree1) = 2\*log m + 3\*log 3 Cost(tree2) = 4\*log m + 5\*log 3

= log m2 + log 33 = log m435

= log m233

Cost(D|tree1) = 7\*log n Cost(D|tree2) = 4\*log n

Cost(tree1,D) = log m233 + 7\*log n Cost(tree2,D) = log m435 + 4\*log n

= log m233n7 = log m435n4

Cost(tree1,D) – Cost(tree2,D)

∴ ควรใช้ tree2 เมื่อ 2n >= 3 + m และควรใช้ tree1 เมื่อ 2n <= 3 + m

5. แบ่งข้อมูลออกเป็น 10 ส่วน แล้วทำการหาความแม่นยำ 10 รอบโดยแต่ละรอบจะใช้ test set 1 ส่วนที่ไม่ซ้ำกัน

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

.

.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |

Train set Test set

ค่าความแม่นยำจะใช้ค่าเฉลี่ยจากทั้ง 10 รอบ และ train set (กล่องสีฟ้า) จะมีการคัดเลือก model จากการแบ่งเป็น 10 ส่วน โดยใช้ 1 ส่วนที่ไม่ซ้ำกัน 10 รอบ เป็น validation set เพื่อหาค่า error และเลือก model ที่มีค่า error รวมต่ำที่สุด

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

.

.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |

Train set validation set