

LAMPIRAN

Perhitungan Algoritma CART, C4.5, dan C5.0 dalam klasifikasi email spam dan ham pada Dataset "data 1 atribut", menggunakan Microsoft Excel dan Google Colab.

1. Lampiran Perhitungan Manual menggunakan Microsoft Excel 2010 :

- Algoritma CART

Tabel Lampiran 1. 1 Hasil Perhitungan Manual Algoritma CART

Node	Spam	Ham	Total	Proporsi Spam	Proporsi Ham	Gini
Root	1.095	747	1.842	0.59	0.40	0.48
Left (≤ 0.1)	1.094	624	1.718	0.63	0.36	0.46
Right (>0.1)	0	127	127	0.0	0	0.0
Setelah Split	-	-	-	-	-	0.43

Reduction in Gini = $0.48 - 0.43 = 0.05$

- Algoritma C4.5

Tabel Lampiran 1. 2 Hasil Perhitungan Manual Algoritma C4.5

Node	Spam	Ham	Total	Proporsi Spam	Proporsi Ham	Entropy	Weight Entropy	IG	Split Info
Root	1.095	747	1.842	0.59	0.40	0.97	-	-	-
Left (≤ 0.1)	1.094	624	1.718	0.63	0.36	0.94	-	-	-
Right (>0.1)	0	127	127	0.0	0.99	0.00	-	-	-
Setelah Split	-	-	-	-	-	0.88	0.88	0.09	0.35

Gain Ratio = $0.9 / 0.35 = 0.25$

- Algoritma C5.0

Tabel Lampiran 1. 3 Hasil Perhitungan Manual Algoritma C5.0

Node	Spam	Ham	Total	Proporsi Spam	Proporsi Ham	Entropy	Weight Entropy	IG	Split Info
Root	1.095	747	1.842	0.59	0.40	0.97	-	-	-
Left ($\leq$ 0.1)	1.094	624	1.718	0.63	0.36	0.94	-	-	-
Right (>0.1)	0	127	127	0.0	0.99	0.00	-	-	-
Setelah Split	-	-	-	-	-	0.87	0.87	0.09	0.35

$$\text{Gain Ratio} = 0.9 / 0.35 = 0.26$$

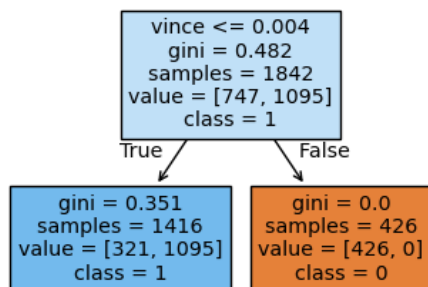
2. Lampiran Hasil Perhitungan menggunakan Google Colab :

- Algoritma CART

```

Indeks Gini sebelum split: 0.4822
Gini Index di node kiri ( $\leq 0.1$ ): 0.4619
Gini Index di node kanan ( $> 0.1$ ): 0.0000
Gini Index setelah split: 0.4303
Reduction in Gini: 0.0519
Threshold yang digunakan model: 0.003746749833226204
{'Gini_before_split': np.float64(0.4821536568027247), 'Gini_left': np.float64(0.46185021272433857)}

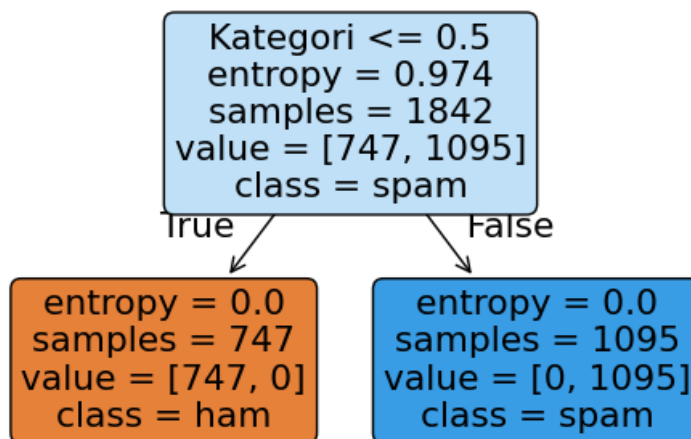
```



Gambar Lampiran 1. 1 Hasil Perhitungan menggunakan Google Colab Algoritma CART

- Algoritma C4.5

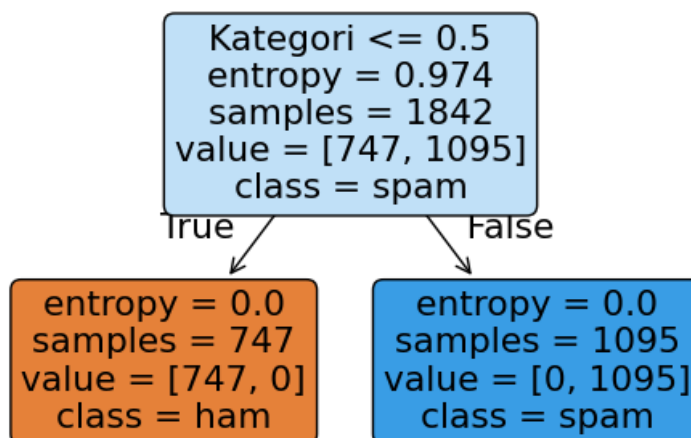
Entropi sebelum split: 0.9741
 Entropi di node kiri (≤ 0.1): 0.9442
 Entropi di node kanan (> 0.1): -0.0000
 Entropi setelah split: 0.8796
 Information Gain: 0.0944
 Split Information: 0.3599
 Gain Ratio: 0.2624
 Threshold yang digunakan model: 0.5



Gambar Lampiran 1. 2 Hasil Perhitungan menggunakan Google Colab Algoritma C4.5

- Algoritma C5.0

Entropi sebelum split: 0.9741
 Entropi di node kiri (≤ 0.1): 0.9442
 Entropi di node kanan (> 0.1): -0.0000
 Entropi setelah split: 0.8796
 Information Gain: 0.0944
 Split Information: 0.3599
 Gain Ratio: 0.2624
 Threshold yang digunakan model: 0.5



Gambar Lampiran 1. 3 Hasil Perhitungan menggunakan Google Colab Algoritma C5.0

3. Perbandingan hasil perhitungan

Tabel Lampiran 2. 1 Perbandingan Hasil Pehitungan dengan dataset 1 atribut

Algoritma	Gini / Entropy Root	Gini / Entropy Setelah Split	Gini / Gain Ratio	Visualisasi Pohon
CART	0.48	0.43	0.05	Ada
C4.5	0.97	0.87	0.26	Ada
C5.0	0.97	0.87	0.26 (dengan pruning	Ada

