

# Laporan Tugas Program Decision Tree

## 1. Pembagian Training dan Testing Data

Data dibagi menggunakan K-Fold Cross-Validation dengan K (banyak fold) berasal dari masukan pengguna.

```
# Split a dataset into k folds
def cross_validation_split(dataset, n_folds):
    dataset_split = list()
    dataset_copy = list(dataset)
    fold_size = int(len(dataset) / n_folds)
    for i in range(n_folds):
        fold = list()
        while len(fold) < fold_size:
            index = randrange(len(dataset_copy))
            fold.append(dataset_copy.pop(index))
        dataset_split.append(fold)
    return dataset_split
```

## 2. Hasil Pengukuran Performa Algoritma

Terdapat 3 ukuran performa yang dilihat yaitu akurasi, presisi, dan recall.

$$Akurasi = \frac{True\ Positive + True\ Negative}{Total}$$

$$Presisi = \frac{True\ Positive}{Actual\ Results} \quad \text{atau} \quad Presisi = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Presisi = \frac{True\ Positive}{Predicted\ Results} \quad \text{atau} \quad Presisi = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Berikut adalah performa untuk penggunaan algoritma dengan 10 Folds Cross-Validation, Kedalaman maksimum tree 5, dan jumlah minimum row pada suatu node 10:

a. Gini Index

[illegible]

N Folds	1	2	3	4	5	6	7	8	9	10
Presisi	95.59 %	95.59 %	95.59 %	95.59 %	95.59 %	95.59 %	95.59 %	95.59 %	95.59 %	95.59 %
Rata-rata	92.86%									

[illegible]

## b. Entropy

[illegible][illegible][illegible]

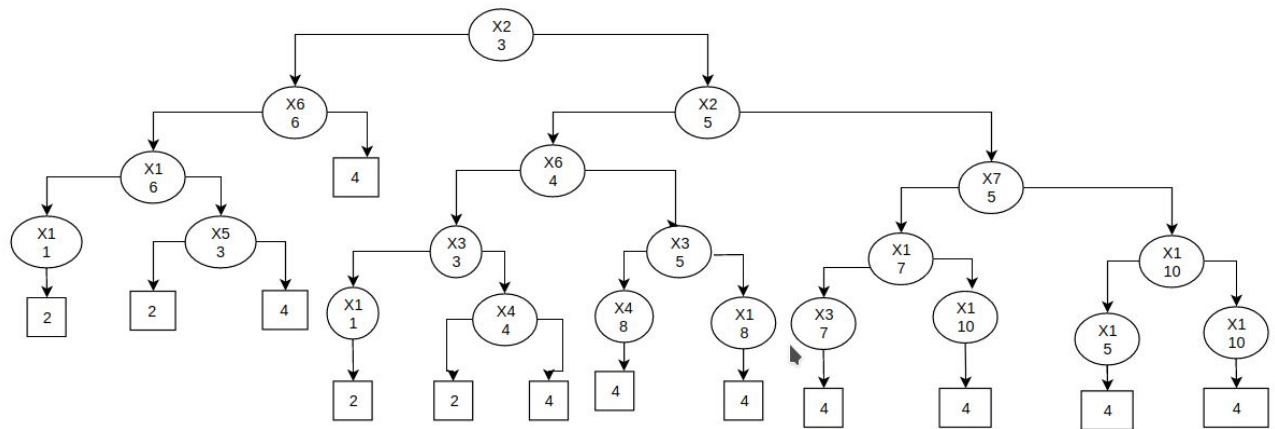
### 3. Model Decision Tree yang Terbentuk

Hasil print di console:

```
Tree:
[X2 < 3.000]
-[X6 < 6.000]
--[X1 < 6.000]
---[X1 < 1.000]
----[2]
----[2]
---[X5 < 3.000]
----[2]
----[4]
--[4]
-[X2 < 5.000]
--[X6 < 4.000]
---[X3 < 3.000]
----[X1 < 1.000]
-----[2]
-----[2]
----[X4 < 4.000]
-----[2]
-----[4]
---[X3 < 5.000]
----[X4 < 8.000]
-----[4]
-----[4]
----[X1 < 8.000]
-----[4]
-----[4]
--[X7 < 5.000]
---[X1 < 7.000]
----[X3 < 7.000]
```

```
-----[4]
-----[4]
----[X1 < 10.000]
-----[4]
-----[4]
---[X1 < 10.000]
----[X1 < 5.000]
-----[4]
-----[4]
----[X1 < 10.000]
-----[4]
-----[4]
```

Jika digambarkan:



Keterangan:

$X_i$  merupakan Kolom ke- $i$

Nilai  $v$  yang menyertai  $X_i$  di setiap node adalah value pembagi. Untuk  $X_i < v$ , ke kiri, sebaliknya ke kanan.

#### 4. Screenshot Run Time Program

```

K untuk Cross-Validation: 10
Kedalaman maximum dari tree: 5
Jumlah row minimum dalam satu node: 10
Rumus Impurity yang digunakan:
1.Entropy
2.Gini
1
melakukan kalkulasi...

Akurasi: =>95.588%, 95.588%, 92.647%, 97.059%, 92.647%, 95.588%, 95.588%, 95.588%, 94.118%, 88.235%,
Rata-rata: 94.265%

Presisi: =>92.857%, 94.737%, 87.500%, 90.476%, 89.655%, 88.000%, 100.000%, 93.103%, 88.000%, 92.308%,
Rata-rata: 91.664%

Recall: =>96.296%, 90.000%, 91.304%, 100.000%, 92.857%, 100.000%, 83.333%, 96.429%, 95.652%, 80.000%,
Rata-rata: 92.587%

Gambar tree untuk akurasi yang terbaik?(1/0)
  
```