TUGAS PERTEMUAN 3

MACHINE LEARNING

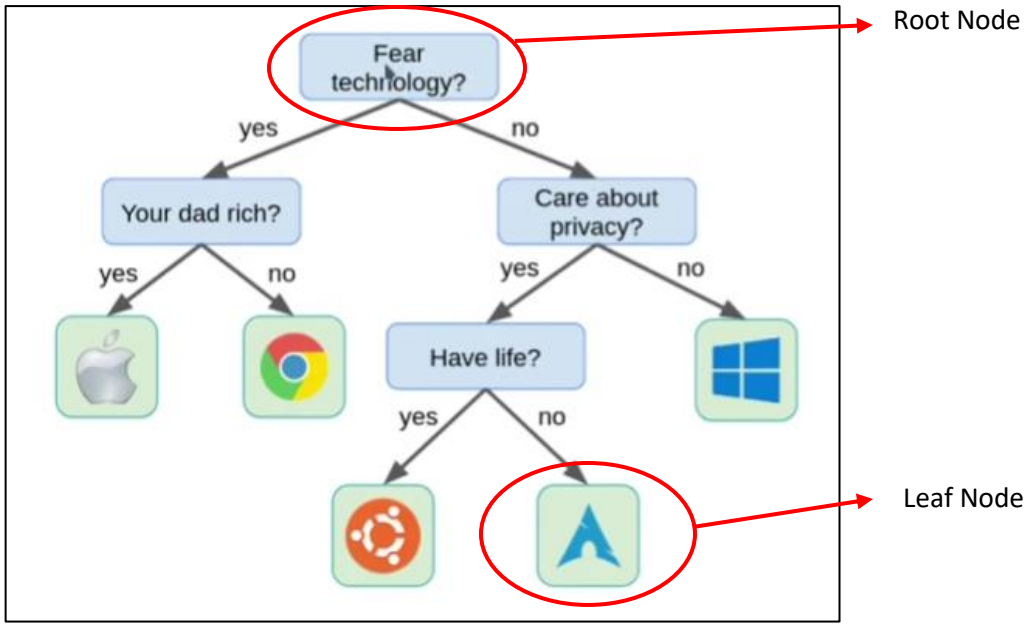
Nama : **Bima Rizki Widiatna**

NPM : **41155050210061**

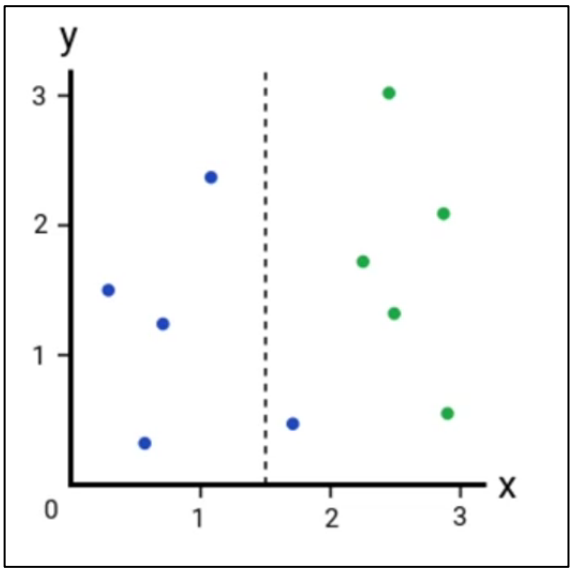
Prodi/Kelas : **Teknik Informatika/A2**

1. ***Decision Tree Classification***

   1.1. Pengenalan Komponen *Decision Tree: Root, Node, Leaf*



1.2. Pengenalan *Gini Impurity*



Gini Impurity memiliki jangkauan nilai antara 0 dan 1, dimana nilai 0 mengindikasikan nilai murni yang sempurna, sedangkan nilai 1 merupakan nilai paling *impure* atau paling tidak murni. Pada gambar di atas terdapat 10 data poin yang terbagi dalam 2 *class*, *class* biru dan *class* hijau.

Pengukuran impurity pada kedua ruas

Ruas Kiri :

$$G = 1 - \sum_{i}^{n} P_i^2$$
$$= 1 - P(biru)^2$$
$$= 1 - (\frac{4}{4})^2 = 0$$

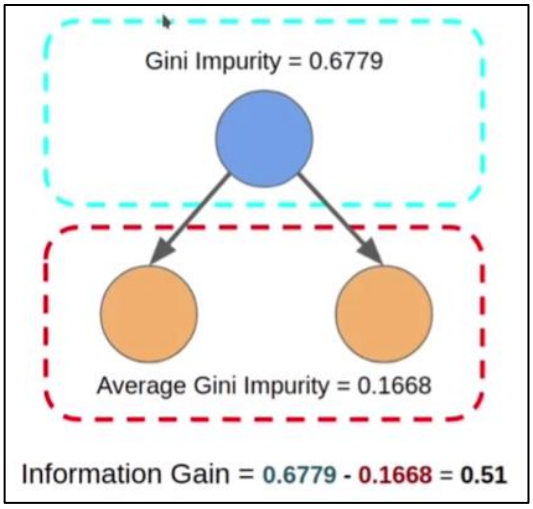
Ruas Kanan :

$$G = 1 - \sum_{i}^{n} P_i^2$$
$$= 1 - (P(biru)^2 + P(hijau)^2)$$
$$= 1 - ((\frac{1}{6})^2 + (\frac{5}{6})^2) = 0.278$$

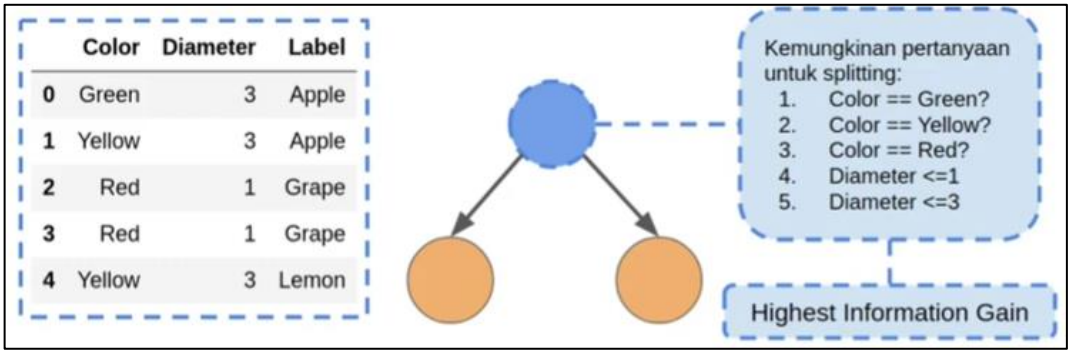
Nilai rata-rata *Gini Impurity*

$$G = \frac{4}{4+6} \times 0 + \frac{6}{4+6} \times 0.278$$
$$= 0.1668$$

1.3. Pengenalan *Information Gain*



1.4. Membangun *Decision Tree*



	Color	Diameter	Label
0	Green	3	Apple
1	Yellow	3	Apple
2	Red	1	Grape
3	Red	1	Grape
4	Yellow	3	Lemon

Pada dataset di atas kolom Color dan Diameter adalah *feature*, dan kolom Label adalah *target*.

Berikut pencarian nilai *Gini Impurity* dari sekumpulan data sebelum dilakukan splitting :

$$G = 1 - (P(apple)^2 + P(grape)^2 + P(lemon)^2)$$
$$= 1 - ((\frac{2}{5})^2 + (\frac{2}{5})^2 + (\frac{1}{5})^2)$$
$$= 0.63$$

1.5. Persiapan *Dataset: Iris Dataset*

[5]:

```
from sklearn.datasets import load_iris

X, y = load_iris(return_X_y=True)

print(f'Dimensi Feature : {X.shape}')
print(f'Class : {set(y)}')
print('\nBima Rizki')
```

```
Dimensi Feature : (150, 4)
Class : {np.int64(0), np.int64(1), np.int64(2)}

Bima Rizki
```

[6]:

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,
                                                    y,
                                                    test_size=0.3,
                                                    random_state=0)
```

1.6. Training Model Decision Tree Classifier

[7]:

```
from sklearn.tree import DecisionTreeClassifier

model = DecisionTreeClassifier(max_depth=4)

model.fit(X_train, y_train)
```

[7]:

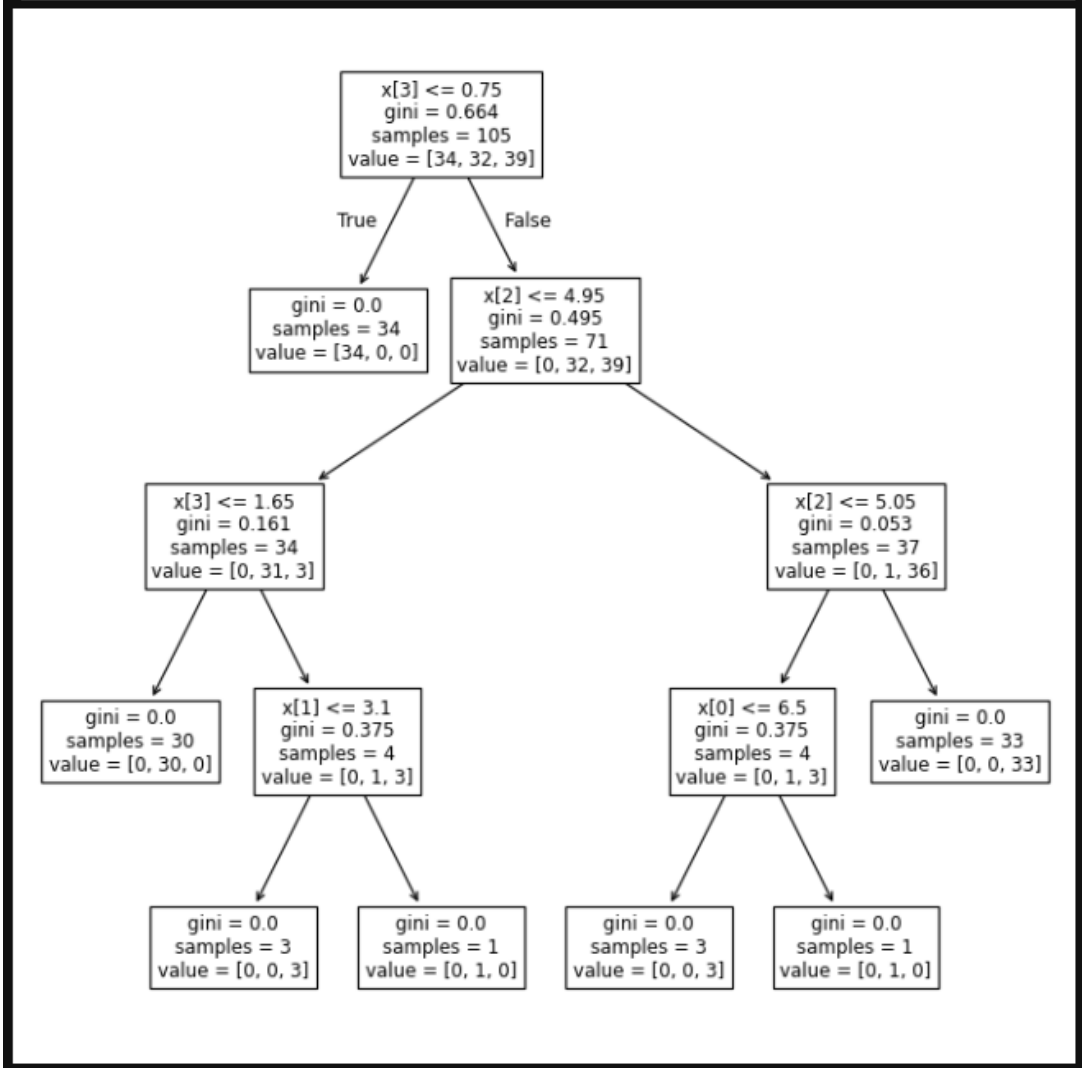
DecisionTreeClassifier

```
DecisionTreeClassifier(max_depth=4)
```

1.7. Visualisasi Model *Decision Tree*

[8]:

```
import matplotlib.pyplot as plt
from sklearn import tree

plt.rcParams['figure.dpi'] = 85
plt.subplots(figsize=(10, 10))
tree.plot_tree(model, fontsize=10)
plt.show()
print('\nBIma Rizki')
```




```
BIma Rizki
```

1.8. Evaluasi Model *Decision Tree*

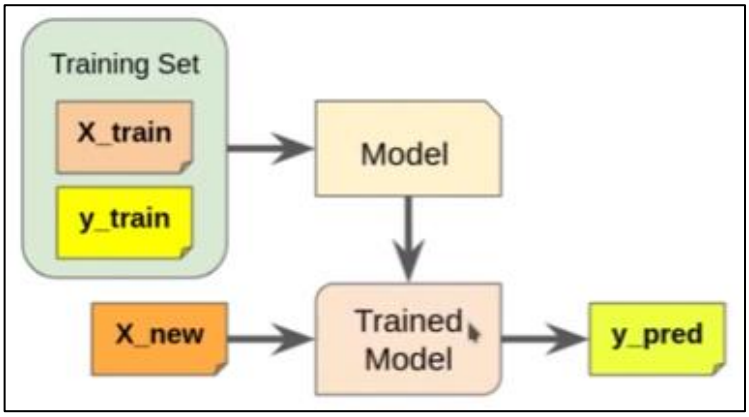
```
[9]:
from sklearn.metrics import classification_report

y_pred = model.predict(X_test)

print(classification_report(y_test, y_pred))
print('\nBima Rizki')
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        16
           1       1.00      0.94      0.97        18
           2       0.92      1.00      0.96        11

    accuracy                           0.98        45
   macro avg       0.97      0.98      0.98        45
weighted avg       0.98      0.98      0.98        45


Bima Rizki
```

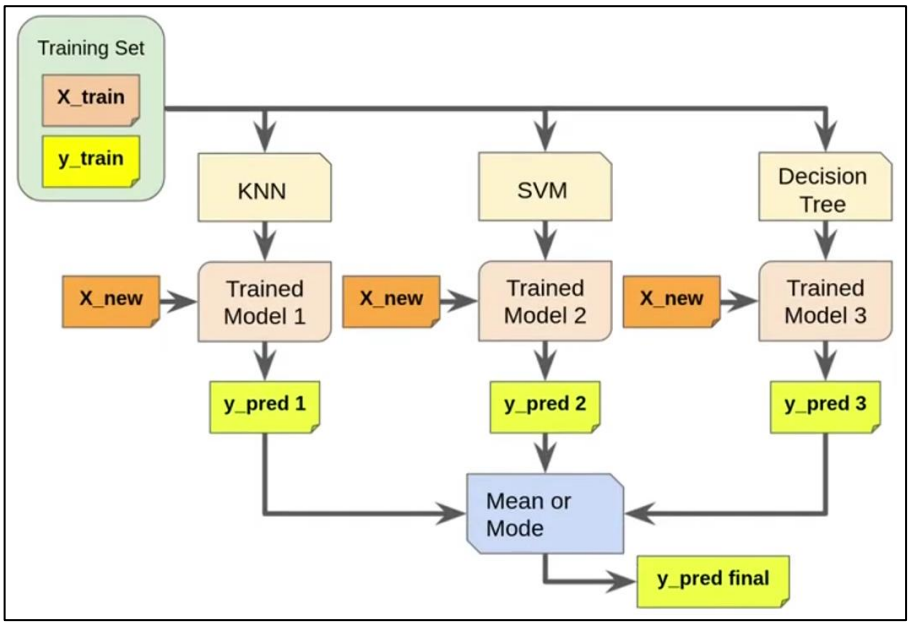
2. *Random Forest Classification*

2.1. Proses Training Model *Machine Learning* Secara Umum

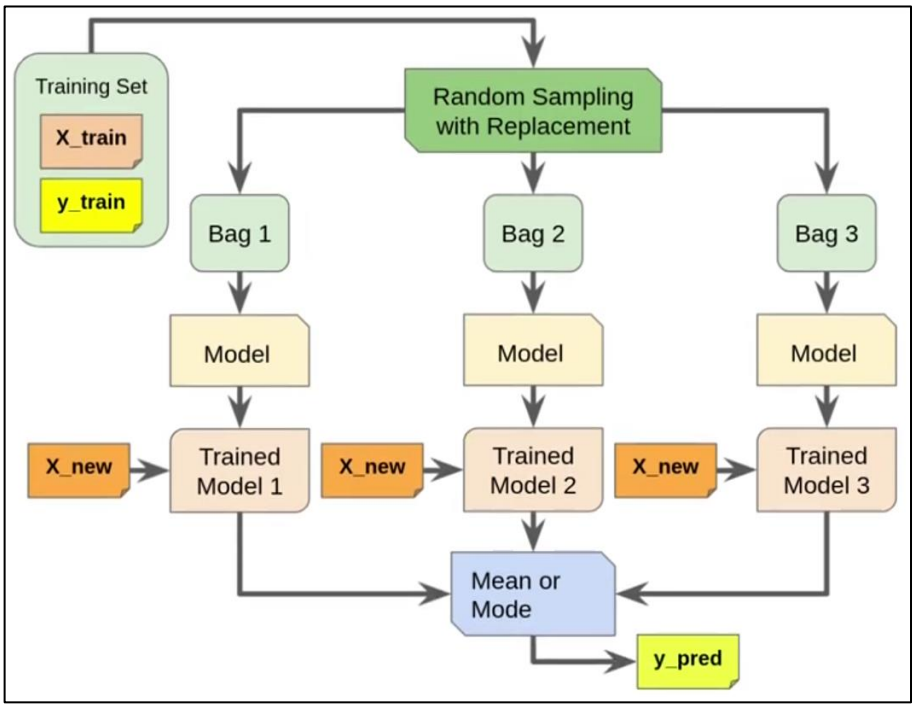
- **General Workflow ML Model Training**



2.2. Pengenalan *Ensemble Learning*

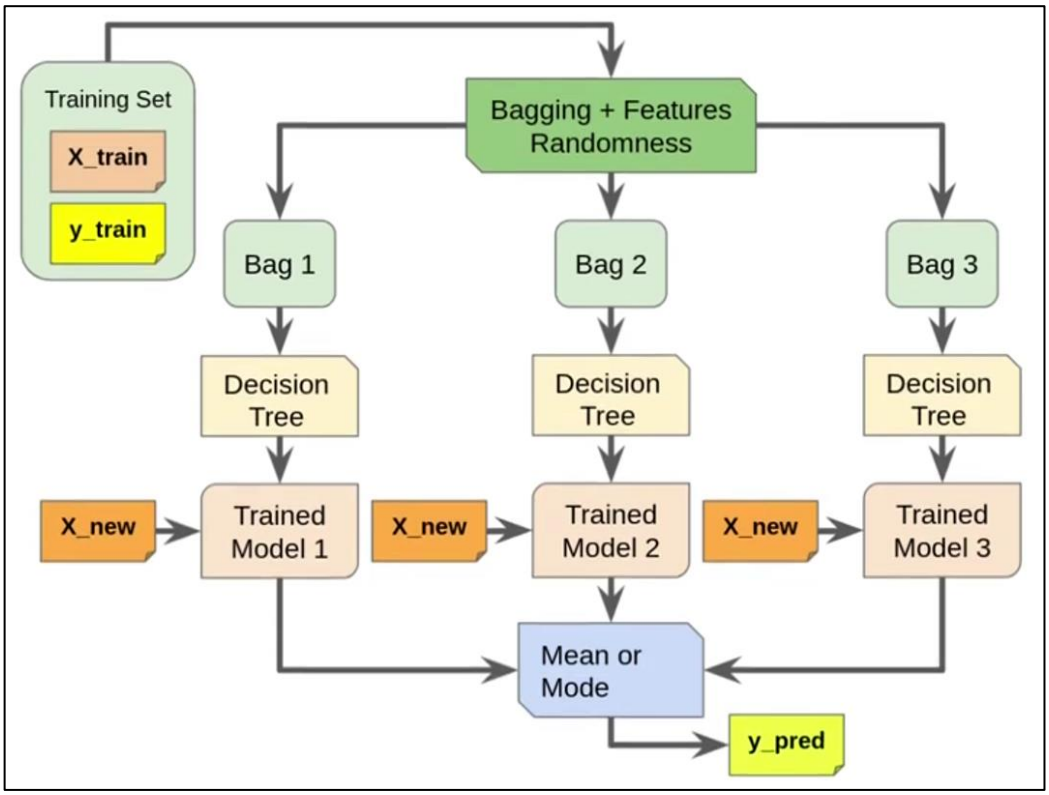


2.3. Pengenalan *Bootstrap Aggregating* | *Bagging*



Dari 3 buah *trained* model pada gambar di atas akan menghasilkan prediksi yang berbeda-beda dari tiap *trained* modelnya. Maka dari itu dari ketiga prediksi tersebut perlu disatukan agar menghasilkan satu prediksi final.

2.4. Pengenalan *Random Forest* | Hutan Acak



Random Forest merupakan implementasi dari *Homogenous Ensemble Learning* yang menerapkan *Decision Tree.*

2.5. Persiapan Dataset | *Iris Flower Dataset*

Dataset

[1]:
```
from sklearn.datasets import load_iris

X, y = load_iris(return_X_y=True)

print(f'Dimensi Feature : {X.shape}')
print(f'Class : {set(y)}')
print('\nBima Rizki')
```

```
Dimensi Feature : (150, 4)
Class : {np.int64(0), np.int64(1), np.int64(2)}

Bima Rizki
```

```
[2]: from sklearn.model_selection import train_test_split

     X_train, X_test, y_train, y_test = train_test_split(X,
                                                         y,
                                                         test_size=0.3,
                                                         random_state=0)
```

2.6. Implementasi *Random Forest Classifier* Dengan *Scikit Learn*

```
[5]: from sklearn.ensemble import RandomForestClassifier

     model = RandomForestClassifier(n_estimators=100,
                                    random_state=0)

     model.fit(X_train, y_train)

[5]: RandomForestClassifier
     RandomForestClassifier(random_state=0)
```

2.7. Evaluasi Model Dengan *Classification Report*

```
[6]: from sklearn.metrics import classification_report

     y_pred = model.predict(X_test)

     print(classification_report(y_test, y_pred))
     print('\nBima Rizki')
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        16
           1       1.00      0.94      0.97        18
           2       0.92      1.00      0.96        11

    accuracy                           0.98        45
   macro avg       0.97      0.98      0.98        45
weighted avg       0.98      0.98      0.98        45


Bima Rizki
```