

## Summary Data Mining

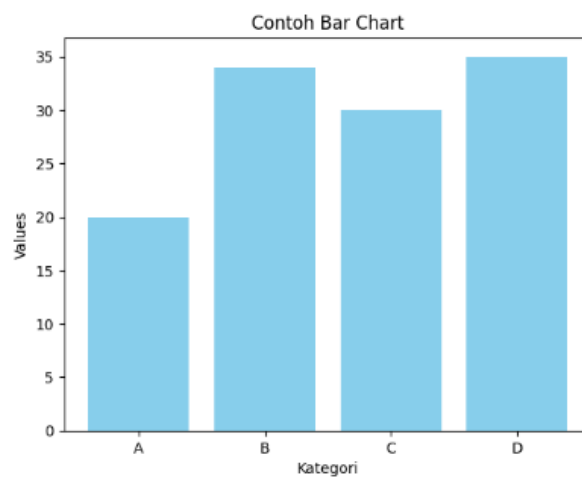
Summary / Ringkasan materi sebelum UTS. pandas dan matplotlib menjadi alat yang sangat berguna dalam melakukan pra-pemrosesan data. Kita belajar bagaimana memuat data ke dalam DataFrame menggunakan pandas, Kemudian, visualisasi data menggunakan matplotlib Dengan membuat berbagai jenis plot seperti histogram, scatter plot, atau bar chart,

```
import pandas as pd
import matplotlib.pyplot as plt

data = {'Kategori': ['A', 'B', 'C', 'D'], 'Values': [20, 34, 30, 35]}
cv = pd.DataFrame(data)

plt.bar(cv['Kategori'], cv['Values'], color='skyblue')
plt.xlabel("Kategori")
plt.ylabel("Values")
plt.title("Contoh")
plt.show()
```

*Gambar. Sintaks yang digunakan*



*Gambar. Output dari Sintaks yang digunakan*

Metode Naive Bayes adalah sekumpulan algoritma pembelajaran terbimbing yang didasarkan pada penerapan teorema Bayes dengan asumsi “naif” tentang independensi bersyarat antara setiap pasangan fitur yang diberikan nilai variabel kelas. Naive Bayes memiliki beberapa jenis utama yaitu Gaussian, Multinomial, dan Bernoulli,

```
from sklearn.datasets import make_classification

X, y = make_classification(
    n_features=6,
    n_classes=3,
    n_samples=800,
    n_informative=2,
    random_state=1,
    n_clusters_per_class=1,
)

import matplotlib.pyplot as plt

plt.scatter(X[:, 0], X[:, 1], c=y, marker="*");
```



```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.33, random_state=125
)

from sklearn.naive_bayes import GaussianNB

model = GaussianNB()

model.fit(X_train, y_train)

predicted = model.predict([X_test[6]])

print("Actual Value:", y_test[6])
print("Predicted Value:", predicted[0])

Actual Value: 0
Predicted Value: 0
```



```
from sklearn.metrics import (
    accuracy_score,
    confusion_matrix,
    ConfusionMatrixDisplay,
    f1_score,
)

y_pred = model.predict(X_test)
accuracy = accuracy_score(y_pred, y_test)
f1 = f1_score(y_pred, y_test, average="weighted")

print("Accuracy:", accuracy)
print("F1 Score:", f1)

Accuracy: 0.8484848484848485
F1 Score: 0.8491119695890328
```

*Gambar: Contoh Sintaks Naïve Bayes GaussianNB serta Outputnya*

Decision Tree, adalah algoritma klasifikasi dan regresi yang berbentuk struktur pohon. Setiap node internal dalam pohon mewakili pengujian pada suatu fitur, setiap cabang mewakili hasil dari pengujian tersebut, dan setiap node daun mewakili label kelas atau nilai prediksi.