

Summary materi Data Mining sebelum UTS

1. Pandas

Pandas itu kayak alat bantu buat ngolah data di Python. Biasanya dipake buat baca data dari Excel atau CSV, terus bisa dimanipulasi—kayak milih kolom tertentu, ngehapus data kosong, atau ngubah format. Intinya, biar ngatur data jadi lebih gampang.

Contoh :

```
import pandas as pd
data = pd.read_csv("data.csv")
print(data.head())
```

Jadi bisa liat data dari file dan diproses dengan gampang.

2. Matplotlib

Matplotlib itu dipake buat bikin grafik. Kalo kamu punya data, terus mau liat bentuknya kayak apa—misal grafik garis, batang, atau pie—nah ini dia alatnya.

Contoh :

```
import matplotlib.pyplot as plt
plt.plot([1,2,3], [4,5,6])
plt.show()
```

Itu bakal nampilin grafik garis dari data yang kita masukan.

3. Naive Bayes

Naive Bayes itu algoritma buat klasifikasi data. Jadi, misal kamu mau nentuin email itu spam atau bukan, algoritma ini bisa bantu. Dia pake logika probabilitas, dan "naive"-nya karena nganggap semua fitur (data input) itu berdiri sendiri padahal di dunia nyata kadang saling nyambung.

Contoh simpel: kalo kata "promo" sering muncul di email spam, maka kalau ada email baru yang ada kata "promo", kemungkinan besar itu spam.

4. Decision Tree

Decision Tree itu kayak pohon keputusan. Jadi datanya dianalisis lewat serangkaian pertanyaan "ya/tidak" sampai ketemu kesimpulan. Misalnya, kalo suhu panas dan hujan, bawa payung. Kalo suhu dingin tapi ga hujan, ga usah.

Visualisasi :

```
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
```

Jadi kalau diibaratkan:

- **Pandas** = buat nyiapin dan beresin data
- **Matplotlib** = buat nampilin data dalam bentuk grafik
- **Naive Bayes & Decision Tree** = buat bikin prediksi atau klasifikasi dari data