Summary, dalam konteks sistem manajemen basis data (DBMS), merujuk pada proses atau teknik untuk menghasilkan data ringkasan dari data yang lebih besar atau kompleks yang disimpan dalam database. Dalam summary meliputi:

1. Pandas: Pengolahan dan Persiapan Data

Pandas adalah library Python utama untuk membaca, mengelola, dan memproses data dalam format tabular (seperti CSV, Excel, dll). Dalam proses analisis data, langkah pertama adalah memuat dataset menggunakan fungsi seperti pd.read\_csv() dan memeriksa isi data menggunakan df.head(), df.info(), atau df.describe().

2. Matplotlib: Visualisasi Data

Matplotlib adalah library visualisasi yang digunakan untuk memahami distribusi dan pola dalam data sebelum pemodelan dilakukan. Contohnya, histogram bisa digunakan untuk melihat distribusi nilai dari satu fitur, sedangkan scatter plot dapat memperlihatkan hubungan antara dua variabel.

3. Naive Bayes: Algoritma Klasifikasi Probabilistik

Naive Bayes adalah algoritma klasifikasi yang sederhana namun sangat efektif, terutama untuk data teks. Ia bekerja berdasarkan Teorema Bayes, yaitu menghitung probabilitas suatu data termasuk ke dalam kelas tertentu, dengan asumsi bahwa setiap fitur bersifat independen.

4. Decision Tree: Algoritma Berbasis Struktur Pohon

Decision Tree adalah algoritma klasifikasi yang membuat model dalam bentuk pohon keputusan, di mana setiap node memisahkan data berdasarkan kondisi pada fitur tertentu. Proses pemisahan ini dilakukan hingga mencapai kondisi yang optimal dalam memisahkan kelas target.