

**Laporan Kelompok**  
**UTS Machine Learning**  
**TI-21-PA**



**Nama Kelompok:**

Frila Cahya Wardani - 212310014

Hera Dwi Pradita - 212310054

Raden Rayyan Pratama Rakhmadie – 212310034

## **Jawaban UTS**

### **1. Identifikasi nama dan jenis atribut dari seluruh atribut prediktor dan atribut label pada dataset**

**Atribut Prediktor:** battery\_power, blue, clock\_speed, dual\_sim, fc, four\_g, int\_memory, m\_dep, mobile\_wt, n\_cores, pc, px\_height, px\_width, ram, sc\_h, sc\_w, talk\_time, three\_g, touch\_screen, dan wifi.

**Atribut Label:** price\_range

### **2. Statistik deskriptif dari data, sebelum dan sesudah preprocessing data**

#### **Sebelum Preprocessing Data:**

Statistik deskriptif memberikan gambaran umum mengenai karakteristik dataset sebelum dilakukannya preprocessing data. Hal tersebut mencakup ukuran seperti rata-rata (mean), nilai tengah (median), nilai maksimum dan minimum, serta penyimpangan standar (standard deviation) untuk setiap fitur.

#### **Setelah Preprocessing Data:**

Statistik deskriptif pada dataset yang sudah dilakukan preprocessing data memberikan gambaran tentang bagaimana preprocessing memengaruhi data. Preprocessing biasanya melibatkan penanganan missing values, normalisasi, standarisasi, atau penghapusan outliers, sehingga data lebih siap untuk analisis atau modeling.

### **3. Model Klasifikasi**

Pada kode program yang kami buat, kami menggunakan model klasifikasi dengan algoritma Decision Tree. Kemudian, pada algoritma tersebut kami juga menggunakan metode holdout untuk membagi dataset menjadi data training dan data testing. Pembagian dataset menjadi data training dan testing dibagi dengan proporsi 85% data training dan 15% data testing. Selain itu, kami menampilkan hasil output dari confusion matrix dari model klasifikasi tersebut dengan nilai sebesar 81% akurat.

#### **4. Model Clustering**

Kami menggunakan model clustering dengan algoritma K-Means, dengan parameter kluster sebesar 5 kluster. Kemudian, dilakukanlah pelatihan untuk model tersebut dan memprediksi nilai kluster. Setelah melatih dan memprediksi nilai kluster, lalu ditampilkan jumlah sample untuk setiap kluster. Setelah itu, kami menampilkan nilai dari silhouette score pada model clustering tersebut. Nilai yang didapatkan ialah sebesar 0.25