

# **LAPORAN HASIL ULANGAN TENGAH SEMESTER**

Disusun untuk memenuhi persyaratan Mata Kuliah Machine Learning Program  
Studi Sarjana Teknologi Informasi

Di Kerjakan Oleh :

**MUHAMMAD ZACKY AFIFF - 212310010**

**MUHAMMAD FAUZAN RINDU MAULANA – 212310011**

**MUHAMMAD FAWAZ TAUHID RIDHO– 212310040**



**PROGRAM STUDI SARJANA TEKNOLOGI INFORMASI  
FAKULTAS INFORMATIKA DAN PARIWISATA  
INSTITUT BISNIS DAN INFORMATIKA KESATUAN  
KOTA BOGOR**

**2024**

**A.** Pembahasan mengenai nama dan jenis atribut dari setiap atribut prediktor dan atribut label yang terdapat pada dataset

- Identifikasi Atribut: Dataset yang digunakan adalah dataset smartphone yang terdiri dari 20 atribut prediktor dan 1 atribut label bernama price\_range. Dataset ini mengandung beberapa nilai yang hilang (missing values) yang perlu ditangani.

Atribut Prediktor:

- battery\_power: Tipe numerik (int64), merepresentasikan daya tahan baterai dalam mAh.
- blue: Tipe kategorikal (int64), menunjukkan ada tidaknya bluetooth (1: Ada, 0: Tidak Ada).
- clock\_speed: Tipe numerik (float64), menunjukkan kecepatan prosesor.
- dual\_sim: Tipe kategorikal (int64), menunjukkan dukungan dual SIM (1: Ada, 0: Tidak Ada).
- fc: Tipe numerik (int64), menunjukkan resolusi kamera depan dalam megapiksel.
- four\_g: Tipe kategorikal (int64), menunjukkan dukungan 4G (1: Ada, 0: Tidak Ada).
- int\_memory: Tipe numerik (int64), menunjukkan memori internal dalam GB.
- m\_dep: Tipe numerik (float64), menunjukkan ketebalan ponsel dalam cm.
- mobile\_wt: Tipe numerik (int64), menunjukkan berat ponsel dalam gram.
- n\_cores: Tipe kategorikal (int64), menunjukkan jumlah core prosesor.
- pc: Tipe numerik (int64), menunjukkan resolusi kamera utama dalam megapiksel.
- px\_height: Tipe numerik (int64), menunjukkan resolusi layar (tinggi) dalam piksel.
- px\_width: Tipe numerik (int64), menunjukkan resolusi layar (lebar) dalam piksel.
- ram: Tipe numerik (int64), menunjukkan RAM dalam MB.
- sc\_h: Tipe numerik (int64), menunjukkan tinggi layar dalam cm.
- sc\_w: Tipe numerik (int64), menunjukkan lebar layar dalam cm.
- talk\_time: Tipe numerik (int64), menunjukkan waktu bicara dalam jam.
- three\_g: Tipe kategorikal (int64), menunjukkan dukungan 3G (1: Ada, 0: Tidak Ada).
- touch\_screen: Tipe kategorikal (int64), menunjukkan ada tidaknya layar sentuh (1: Ada, 0: Tidak Ada).
- wifi: Tipe kategorikal (int64), menunjukkan ada tidaknya WiFi (1: Ada, 0: Tidak Ada).

- Atribut Label:
  - `price_range`: Tipe kategorikal (`int64`), merepresentasikan rentang harga ponsel, dengan nilai unik: 0 (Low Cost), 1 (Medium Cost), 2 (High Cost), 3 (Very High Cost).

**B.** Pembahasan mengenai statistik deskriptif dari data, baik untuk data sebelum dilakukan praproses maupun data setelah dilakukan pengisian missing values dan standarisasi

1. Sebelum Praproses:

- a. Fungsi `df.describe()` digunakan untuk menampilkan statistik deskriptif awal data, seperti rata-rata, standar deviasi, nilai minimum, kuartil, dan nilai maksimum untuk setiap atribut numerik.
- b. Fungsi `null_count()` didefinisikan untuk menghitung dan menampilkan jumlah dan persentase *missing values* pada setiap atribut.

2. Setelah Praproses:

- a. `SimpleImputer` digunakan untuk mengisi *missing values* dengan nilai rata-rata atribut.
- b. `StandardScaler` digunakan untuk standarisasi data, membuat nilai atribut memiliki rata-rata 0 dan standar deviasi 1.
- c. `df.describe()` digunakan kembali untuk menampilkan statistik deskriptif setelah praproses. Perhatikan perubahan pada nilai rata-rata dan standar deviasi setelah standarisasi.

**C.** Pembahasan mengenai model klasifikasi

Model Klasifikasi

1. Algoritma: Support Vector Machine (SVM) digunakan untuk membangun model klasifikasi.
2. Pembagian Data: Metode *holdout* dengan `train_test_split` digunakan untuk membagi data menjadi data latih (85%) dan data uji (15%).
3. Evaluasi:
  - a. `confusion_matrix` digunakan untuk membuat matriks konfusi, yang menunjukkan kinerja model dalam mengklasifikasikan setiap kelas.
  - b. `accuracy_score` digunakan untuk menghitung akurasi model, yaitu persentase prediksi yang benar.
  - c. `classification_report` digunakan untuk menampilkan laporan klasifikasi yang mencakup *precision*, *recall*, \*

**D. Pembahasan mengenai model clustering**

1. Kode yang diberikan mengimplementasikan model clustering menggunakan algoritma K-Means.
2. Data distandarisasi terlebih dahulu sebelum dilakukan clustering.
3. Elbow Method digunakan untuk menentukan jumlah cluster optimal, yaitu 2 cluster.
4. Silhouette Score digunakan untuk mengevaluasi kualitas cluster, dengan skor yang diperoleh sebesar [Silhouette Score].
5. Hasil clustering divisualisasikan menggunakan scatter plot.

SOURCE CODE :

[https://drive.google.com/drive/folders/1Qlt5nFblP\\_3uHvC7qcmbeXdugpGt-6l?usp=sharing](https://drive.google.com/drive/folders/1Qlt5nFblP_3uHvC7qcmbeXdugpGt-6l?usp=sharing)