

LAPORAN UTS MACHINE LEARNING



Disusun oleh:

Erdiana Ragil Syawala – 212310045 – TI21PA

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

2024

A. Pembahasan Mengenai Nama dan Jenis Atribut dari Setiap Atribut Predictor dan atribut Label

Dalam dataset yang diberikan, terdapat 20 atribut predictor dan 1 atribut label. Berikut adalah nama dan jenis atribut beserta jumlah nilai unik untuk masing-masing atribut:

Atribut Predictor:

- **battery_power:** Numerik, 1093 nilai unik.
- **blue:** Kategorik, 2 nilai unik (0, 1).
- **clock_speed:** Numerik, 26 nilai unik.
- **dual_sim:** Kategorik, 2 nilai unik (0, 1).
- **fc:** Numerik, 20 nilai unik.
- **four_g:** Kategorik, 2 nilai unik (0, 1).
- **int_memory:** Numerik, 63 nilai unik.
- **m_dep:** Kategorik, 10 nilai unik (contoh: 0.6, 0.7, 0.9, dll).
- **mobile_wt:** Numerik, 121 nilai unik.
- **n_cores:** Kategorik, 8 nilai unik (contoh: 1, 2, 3, dll).
- **pc:** Numerik, 21 nilai unik.
- **px_height:** Numerik, 1137 nilai unik.
- **px_width:** Numerik, 1109 nilai unik.
- **ram:** Numerik, 1558 nilai unik.
- **sc_h:** Numerik, 15 nilai unik.
- **sc_w:** Numerik, 19 nilai unik.
- **talk_time:** Numerik, 19 nilai unik.
- **three_g:** Kategorik, 2 nilai unik (0, 1).
- **touch_screen:** Kategorik, 2 nilai unik (0, 1).
- **wifi:** Kategorik, 2 nilai unik (1, 0).

Atribut Label:

- **price_range:** Kategorik, 4 nilai unik (0, 1, 2, 3).

B. Pembahasan Mengenai Statistik Deskriptif dari Data

Sebelum Praproses:

Statistik deskriptif pada dataset sebelum dilakukan praproses (penanganan missing values dan standarisasi) menunjukkan informasi dasar mengenai data, seperti rata-rata, nilai minimum dan maksimum, serta variasi data untuk setiap atribut. Beberapa contoh statistik deskriptif dari atribut numerik adalah sebagai berikut:

- **battery_power:** Rata-rata = 1237.87, Min = 501, Max = 1998.
- **blue:** Rata-rata = 0.495, dengan nilai unik 0 dan 1.
- **clock_speed:** Rata-rata = 1.52, Min = 0.5, Max = 3.
- **fc:** Rata-rata = 4.31, Min = 0, Max = 19.
- **mobile_wt:** Rata-rata = 140.34, Min = 80, Max = 200.
- **ram:** Rata-rata = 2124.99, Min = 256, Max = 3998.

Setelah Praproses:

Setelah pengisian nilai hilang (missing values) menggunakan SimpleImputer dengan strategi rata-rata dan standarisasi data menggunakan StandardScaler, atribut-atribut numerik diubah menjadi standar yang memiliki nilai rata-rata mendekati 0 dan standar deviasi mendekati 1. Beberapa contoh statistik setelah praproses adalah sebagai berikut:

- **battery_power:** Mean = -1.56e-16, Std = 1.00.
- **blue:** Mean = -1.24e-17, Std = 1.00.

Praproses ini memastikan bahwa seluruh data berada dalam skala yang konsisten dan siap untuk digunakan dalam model klasifikasi dan clustering.

C. Pembahasan Mengenai Model Klasifikasi

Pada bagian ini, model klasifikasi dibangun menggunakan **Decision Tree** (atau bisa menggunakan SVM sebagai alternatif). Dataset dibagi menjadi 85% untuk data training dan 15% untuk data testing menggunakan metode **holdout**.

Setelah melakukan pelatihan dengan algoritma **Decision Tree**, model diuji menggunakan data testing, dan hasil evaluasi model berupa **confusion matrix** dan **akurasi** dihitung. Confusion matrix menunjukkan seberapa baik model dapat mengklasifikasikan data berdasarkan prediksi yang benar dan salah, sedangkan akurasi menunjukkan persentase prediksi yang benar dari total data.

D. Pembahasan Mengenai Model Clustering

Pada bagian ini, model clustering dibangun menggunakan **K-Means** untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan fitur. Hasil evaluasi kualitas clustering dilakukan dengan menghitung **silhouette score**, yang menunjukkan seberapa baik data dikelompokkan dalam cluster yang sesuai.

Silhouette score memberikan gambaran mengenai seberapa terpisah dan rapatnya cluster yang terbentuk, dan nilai yang lebih tinggi menunjukkan kualitas clustering yang lebih baik.

Kesimpulan

Berdasarkan hasil analisis, beberapa kesimpulan yang dapat diambil antara lain:

- Praproses data, termasuk pengisian missing values dan standarisasi, sangat penting untuk menyiapkan data yang konsisten dan siap digunakan dalam model.
- Model klasifikasi, seperti **Decision Tree**, memberikan hasil yang cukup baik dengan akurasi yang memadai dalam memprediksi kelas harga smartphone.
- Evaluasi menggunakan **silhouette score** pada model **K-Means** memberikan indikasi bahwa clustering dilakukan dengan baik, meskipun dapat ditingkatkan lebih lanjut untuk kualitas clustering yang lebih optimal.