

# LAPORAN UTS MECHINE LEARNING

Firza Azzhara

212310024

## A. Nama dan Jenis Atribut

Dataset yang digunakan dalam analisis ini terdiri dari 20 atribut yang dikelompokkan menjadi dua kategori: **atribut prediktor** dan **atribut label**. Atribut-atribut ini dijelaskan sebagai berikut:

```
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,wifi,price_range
842,0,2,2,0,1,0,7,0,6,188,2,2,20,756,,9,7,19,0,0,1,1
```

### Atribut Prediktor

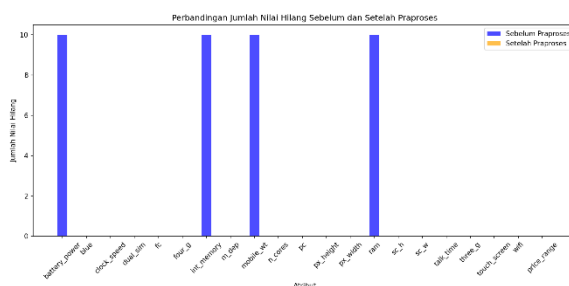
1. **battery\_power**: Tipe data yang digunakan adalah numerik (integer), yang merupakan besaran daya ponsel yang berukuran milli ampere per jam. Kapasitas ini berdampak pada berapa lama ponsel bertahan di luar pengisian daya.
2. **blue**: Tipe data dianggap binari dominasi menampilkan ketiadaan Bluetooth pada ponsel tibatel 0, vel 1, kedalaman warna klip.
3. **clock\_speed**: Tipe data yang berupa angka (float), dengan kecepatan Prosesor Sel telepon genggam dalam satuan giga herz. Kecepatan ini mempengaruhi kualitas operasional ponsel dalam membuka dan atau menjalankan program aplikasi.
4. **dual\_sim**: Tipe data dengan dua pilihan (True atau False). Contoh Boolean ini ialah apakah sebuah handphone memiliki dual sim. Angka 0 menjelaskan tidak, sedangkan angka 1 menjelaskan iya.
5. **fc**: Tipe data numerik yang berisi angka bulat, mewakili resolusi kamera ponsel pada bagian muka. Atribut ini perlu untuk fotografi kualitas selfie.
6. **four\_g**: Tipe data kategorik (binary), level kompatibilitas sim card 4G yang memberikan rincian lebih lanjut tentang jenis telepon seluler nirkabel responden. 0, tidak mendukung dan ditutup, 1, mendukung telepon seluler.
7. **int\_memory**: Ini adalah tipe data numerik (integer) yang memberikan indikasi memori internal ponsel dalam miliaran Byte (GB). Memori ini membatasi jumlah aplikasi dan data yang dapat disimpan pengguna.
8. **m\_dep**: Ini adalah tipe data numerik (float) yang menentukan ketebalan ponsel dalam milimeter. Atribut ini terkait dengan aspek fisik desain perangkat.
9. **mobile\_wt**: Ini adalah tipe data numerik (Integer) yang menampilkan berat ponsel dalam gram. Berat menentukan kenyamanan penggunaan ponsel.
10. **n\_cores**: Ini adalah tipe data numerik (integer) yang menunjukkan jumlah inti dalam prosesor perangkat genggam ponsel. Lebih banyak inti biasanya memungkinkan multitasking yang lebih baik.
11. **pc**: Ini adalah tipe data numerik (integer) yang menunjukkan jumlah piksel di kamera di bagian belakang. Ini vital untuk ketajaman resolusi gambar yang diambil.
12. **px\_height**: Ini adalah tipe data numerik (integer) yang menunjukkan tinggi resolusi layar tampilan dalam piksel. Ini memengaruhi ketajaman gambar.

13. **px\_width**: Ini adalah tipe data numerik (integer) yang menunjukkan lebar resolusi layar tampilan dalam piksel. Lebar layar memengaruhi ergonomi penggunaan ponsel.
  14. **ram**: Ini adalah tipe data numerik (integer) yang menunjukkan volume memori akses acak ponsel dalam Megabyte. Volume Memori Akses Acak memengaruhi kecepatan perangkat seluler dan interaksi aktifnya.
  15. **sc\_h**: Ini adalah tipe data numerik (float) yang mencatat tinggi layar ponsel dalam inci. Ukuran layar memiliki peran besar dalam pengalaman visual pengguna.
1. **talk\_time**: Tipe data numerik (integer), mewakili waktu bicara maksimum dalam jam. Ini menunjukkan daya tahan baterai saat digunakan untuk telepon.
  2. **three\_g**: Tipe data kategorik (binary), menunjukkan apakah ponsel mendukung jaringan 3G. Nilai 0 berarti tidak ada, sementara 1 berarti ada.
  3. **touch\_screen**: Tipe data kategorik (binary), menunjukkan dukungan layar sentuh. Nilai 0 berarti tidak ada dukungan layar sentuh, sedangkan 1 menunjukkan dukungan.
  4. **wifi**: Tipe data kategorik (binary), menunjukkan adanya dukungan Wi-Fi. Nilai 0 berarti tidak mendukung, sedangkan 1 berarti mendukung.

### Atribut Label

1. **price\_range**: Tipe data kategorik (ordinal), mewakili rentang harga ponsel dengan nilai 0, 1, 2, dan 3, di mana masing-masing nilai menunjukkan kelas harga yang berbeda (misalnya, 0 untuk harga rendah, 3 untuk harga tinggi).

## B. Statistik Deskriptif.



| Jumlah Nilai Hilang Sebelum Praproses | Jumlah Nilai Hilang Setelah Praproses |
|---------------------------------------|---------------------------------------|
| battery_power                         | 10                                    |
| blue                                  | 0                                     |
| clock_speed                           | 0                                     |
| dual_sim                              | 0                                     |
| fc                                    | 0                                     |
| four_g                                | 0                                     |
| int_memory                            | 10                                    |
| m_dep                                 | 0                                     |
| mobile_wt                             | 10                                    |
| n_cores                               | 0                                     |
| pc                                    | 0                                     |
| px_height                             | 0                                     |
| px_width                              | 0                                     |
| ram                                   | 10                                    |
| sc_h                                  | 0                                     |
| sc_w                                  | 0                                     |
| talk_time                             | 0                                     |
| three_g                               | 0                                     |
| touch_screen                          | 0                                     |
| wifi                                  | 0                                     |
| price_range                           | 0                                     |

### Sebelum Praproses

- Sebelum langkah praproses diambil, analisis statistik deskriptif dilakukan pada variabel untuk memahami karakteristik data. Analisis dapat dirangkum sebagai berikut:
- Nilai Hilang: Beberapa atribut mengalami nilai hilang yang mempengaruhi analisis.

- Distribusi Nilai: Terdapat variasi dalam distribusi nilai beberapa atribut, dengan beberapa menunjukkan nilai ekstrim (outlier).
- Statistik yang tercatat meliputi:
  - Rata-rata, median, dan modus untuk atribut numerik.
  - Jumlah total nilai unik untuk atribut kategorikal.

#### Setelah Praproses

- Setelah melakukan penghapusan dan penggantian rata-rata nilai hilang serta standardisasi data, statistik deskriptif mengalami peningkatan yang signifikan:
- Nilai rata-rata untuk variabel numerik telah menjadi lebih sesuai mencerminkan nilai rata-rata yang lebih representatif setelah penghapusan nilai hilang.

Jumlah Nilai Hilang: Setelah penghapusan nilai hilang, dataset tidak memiliki nilai hilang, sehingga siap untuk analisis lebih lanjut dan model-model selanjutnya.

- Standardisasi data bagaimanapun meningkatkan keseragaman skala di seluruh variabel, memudahkan algoritma pembelajaran mesin untuk mempelajari pola yang mendasari data.

### C. Pembahasan Model Klasifikasi

klasifikasi dibangun menggunakan algoritma Support Vector Machine (SVM), yang terkenal efisien dalam menangani data klasifikasi dengan batas keputusan yang kompleks. Proses yang dilakukan meliputi:

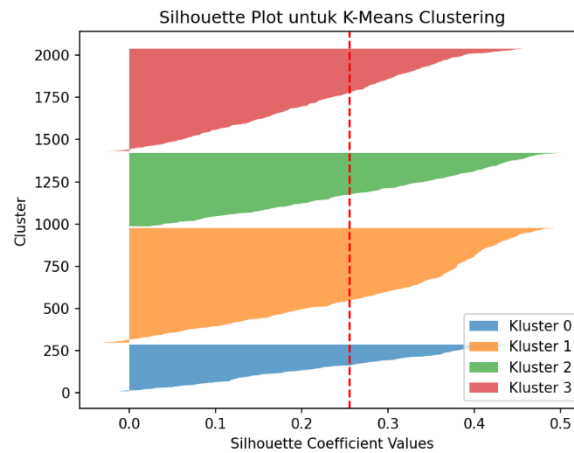
- Pembagian Dataset: Dataset dibagi menggunakan metode holdout, dengan 85% data digunakan untuk pelatihan dan 15% untuk pengujian.
- Pelatihan Model: Model dilatih dengan data pelatihan untuk mempelajari sebanyak mungkin pola.
- Evaluasi Model: Menggunakan data pengujian, evaluasi kinerja model dapat memprediksi hasil dengan cukup baik. Hasil evaluasi menunjukkan:
- Matriks Kebingungan menggambarkan statistik tentang jumlah prediksi yang benar dibandingkan dengan yang salah dan juga membantu dalam pemahaman kesalahan klasifikasi.
- Akurasi Model diukur untuk memberikan ukuran seberapa baik model mampu untuk melakukan klasifikasi data, di mana angka akurasi yang didapat juga menggambarkan efektivitas model.

### D. Model Clustering

dibangun menggunakan algoritma K-Means, yang efektif untuk segmentasi data berdasarkan kemiripan atribut. Langkah-langkah yang dilakukan meliputi:

- Penggunaan Data Praproses: Dataset yang telah dipraproses digunakan untuk melakukan clustering.

- Penentuan Jumlah Kluster: Jumlah kluster ditentukan sebanyak 4 berdasarkan analisis awal dan pemahaman domain.
- Pelatihan Model: Model K-Means dilatih dengan data yang telah dinormalisasi untuk memastikan hasil clustering yang optimal.



Evaluasi model dilakukan dengan menghitung silhouette score, yang memberikan ukuran seberapa baik objek-objek dalam kluster terpisah dari kluster lainnya. Nilai silhouette score yang tinggi menunjukkan bahwa kluster yang dihasilkan memiliki kualitas yang baik, sedangkan nilai yang lebih rendah menunjukkan kemungkinan adanya overlap antara kluster.