

# **LAPORAN JAWABAN UTS (ULANGAN TENGAH SEMESTER)**

## **MACHINE LEARNING**



Emanuel Charel Alessandro Soge 212310041

Muhamad Yannuar Maulana 212310023

Melani 2123100

**PROGRAM STUDI TEKNOLOGI INFORMASI**

**INSTITUT BISNIS DAN INFORMATIKA KESATUAN**

**BOGOR 2024**

## **KATA PENGANTAR**

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat, karunia, dan kesempatan yang diberikan kepada saya sehingga dapat menyelesaikan laporan ini dengan baik. Laporan ini disusun sebagai bagian dari tugas yang diberikan dalam mata pelajaran machine learning.

Dalam proses penyusunan laporan ini, penulis telah berusaha sebaik mungkin untuk menyusun pembahasan. Meskipun demikian, penulis menyadari bahwa masih banyak kekurangan dalam laporan ini. Semoga laporan ini dapat memberikan manfaat bagi pembaca dan dapat memenuhi tujuan yang diharapkan.

### a. Nama dan Jenis Atribut dari Setiap Atribut Prediktor dan Atribut Label

- Atribut Prediktor: Terdiri dari fitur-fitur yang digunakan untuk memprediksi nilai output. Dalam kode, atribut prediktor meliputi `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`. Jenis atributnya ada yang **numerik (continuous)** dan **kategorik (binary)**.
- Atribut Label: Tidak disebutkan secara eksplisit dalam potongan kode, tetapi ini biasanya merupakan variabel target yang diprediksi oleh model.

### b. Statistik Deskriptif Data

- Sebelum Praproses: Statistik deskriptif awal memberikan gambaran seperti mean, standar deviasi, minimum, dan maksimum dari setiap atribut, yang dapat membantu memahami distribusi awal data.
- Setelah Praproses (Pengisian Missing Values dan Standarisasi): Setelah praproses, data lebih bersih dan terstandarisasi, sehingga nilai-nilai berada dalam rentang yang seragam. Ini membantu meningkatkan performa model dengan mengurangi pengaruh dari skala yang berbeda pada atribut.

### c. Model Klasifikasi

- Kode menggunakan Support Vector Machine (SVM) dengan metode *holdout* (85% data untuk training dan 15% data untuk testing). SVM adalah model klasifikasi yang bekerja dengan mencari hyperplane optimal yang memisahkan kelas-kelas dalam data. Visualisasi boundary decision dibuat dengan menggunakan dua fitur utama untuk menggambarkan area prediksi dari SVM.
- Pada bagian lain, juga ada Decision Tree yang menghasilkan *confusion matrix* dan mengukur akurasi untuk menunjukkan kinerja klasifikasi, dengan hasil akurasi sekitar 85%.

#### **d. Model Clustering**

- Kode menggunakan algoritma *K-Means Clustering* untuk mengelompokkan data. K-Means mengelompokkan data berdasarkan jarak terdekat ke pusat klaster (*centroid*).
- *Silhouette Score* digunakan untuk menilai kualitas klaster, dan *Elbow Method* membantu menentukan jumlah klaster optimal. Dari hasil, jumlah klaster optimal ditemukan pada 2 klaster, yang menunjukkan bahwa data dapat dikelompokkan dengan baik ke dalam dua grup.

Berikut link git-hub

<https://github.com/Charelas/UTS-Machine-Learning-24-25>