

UJIAN TENGAH SEMESTER (UTS)- SEMESTER GANJIL

TAHUN AKADEMIK 2025/2026

Mata Kuliah : Machine Learning / TIFA4F3

Semester / Kelas : 7 (Tujuh)/TI-22-PA

Program Studi : S1 Teknologi Informasi

Dosen Pengampu : Isnain Mulia, S.Komp., M.Kom./

Hari, Tanggal : Rabu, 05 November 2025

Waktu Ujian : 16.00-17.30WIB

Sifat Ujian : Closed Book/Open All Prosess

Nama Anggota Kelompok:

Reyhan Javier Rachman_(222310011)_TI-22-PA

Muhammad Faadihilah_(222310021)_TI-22-PA

Fikri Maulana Iskandar _ (222310027) _ TI-22-PA

2.) Mengidentifikasi nama dan jenis atribut dari seluruh atribut prediktor dan atribut label yang terdapat pada dataset tersebut. Sebelum menentukan jenis atribut, Anda dapat memeriksa terlebih dahulu berapa banyak nilai unik yang terdapat di setiap atribut. Untuk atribut yang berjenis kategorik, sebutkan nilai unik yang terdapat pada atribut tersebut.

No	Nama Atribut	Peran	Jenis Atribut	Jumlah Nilai Unik	Nilai Unik (untuk atribut kategorik)
1	Age	Prediktor	Numerik (kontinu/diskrit)	50	-
2	Sex	Prediktor	Kategorik nominal	2	F, M
3	ChestPainType	Prediktor	Kategorik nominal	4	ASY, ATA, NAP, TA
4	RestingBP	Prediktor	Numerik (kontinu/diskrit)	67	-
5	Cholesterol	Prediktor	Numerik (kontinu/diskrit)	219	-
6	FastingBS	Prediktor	Kategorik biner	2	0, 1
7	RestingECG	Prediktor	Kategorik nominal	3	LVH, Normal, ST
8	MaxHR	Prediktor	Numerik (kontinu/diskrit)	119	-
9	ExerciseAngina	Prediktor	Kategorik nominal	2	N, Y
10	Oldpeak	Prediktor	Numerik (kontinu/diskrit)	53	-
11	ST_Slope	Prediktor	Kategorik nominal	3	Down, Flat, Up
12	HeartDisease	Label (Target)	Kategorik biner	2	0, 1

6.) [25 poin] Merangkum apa saja yang dilakukan pada poin 2 s.d. 5 di atas dalam bentuk laporan tertulis. Detail yang harus dituliskan adalah sebagai berikut.

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

b. Pembahasan mengenai statistik deskriptif dari data, sebelum dan setelah dilakukan praproses data.

Jika ada tambahan praproses data yang dilakukan, berikan penjelasannya

c. Pembahasan mengenai model klasifikasi

d. Pembahasan mengenai model clustering

e. Lengkapi laporan dengan nama dan NPM setiap anggota, serta kelas

No	Nama Atribut	Jenis Atribut	Keterangan
1	Age	Numerik (kontinu)	Usia responden
2	Sex	Kategorik	Jenis kelamin (M/F)
3	ChestPainType	Kategorik	Jenis nyeri dada
4	RestingBP	Numerik (kontinu)	Tekanan darah saat istirahat
5	Cholesterol	Numerik (kontinu)	Kadar kolesterol
6	FastingBS	Numerik (biner)	Gula darah puasa (0/1)
7	RestingECG	Kategorik	Hasil pemeriksaan ECG
8	MaxHR	Numerik (kontinu)	Detak jantung maksimum
9	ExerciseAngina	Kategorik	Angina saat olahraga (Y/N)
10	Oldpeak	Numerik (kontinu)	Depresi ST relatif terhadap istirahat
11	ST_Slope	Kategorik	Kemiringan segmen ST
12	HeartDisease	Label (biner)	0 = Tidak, 1 = Mengidap penyakit jantung

- A. Dataset yang digunakan dalam analisis ini terdiri dari dua belas atribut, yaitu sebelas atribut prediktor dan satu atribut label bernama HeartDisease. Atribut prediktor mencakup data numerik seperti Age, RestingBP, Cholesterol, FastingBS, MaxHR, dan Oldpeak, serta atribut kategorik seperti Sex, ChestPainType, RestingECG, ExerciseAngina, dan ST_Slope. Sementara itu, atribut HeartDisease merupakan label dengan nilai biner (0 untuk tidak memiliki penyakit jantung dan 1 untuk memiliki penyakit jantung).

- B. Sebelum dilakukan praproses, data memiliki skala yang bervariasi antar atribut dan terdapat beberapa nilai kosong atau tidak wajar, seperti nilai nol pada kolom RestingBP dan Cholesterol. Oleh karena itu, langkah praproses dilakukan agar data siap digunakan dalam pemodelan. Nilai hilang pada atribut numerik diimputasi menggunakan metode rata-rata (SimpleImputer – mean strategy). Selanjutnya, seluruh atribut numerik distandardisasi menggunakan StandardScaler agar memiliki rentang skala yang seragam (mean = 0 dan standar deviasi = 1). Atribut kategorik diubah menjadi data numerik dengan teknik One-Hot Encoding. Setelah seluruh proses ini, data numerik dan kategorik digabung menjadi satu dataset yang sudah bersih dan siap dianalisis.
- C. Untuk tahap klasifikasi, algoritma yang digunakan adalah Logistic Regression karena dataset ini bersifat biner dan metode tersebut sederhana namun efektif. Data dibagi menjadi 85% untuk data latih dan 15% untuk data uji. Hasil evaluasi menunjukkan bahwa model menghasilkan akurasi sebesar 89,13% dengan confusion matrix [[54, 8], [7, 69]]. Nilai ini menunjukkan bahwa model dapat mengklasifikasikan data pasien dengan dan tanpa penyakit jantung dengan tingkat ketepatan yang cukup tinggi
- D. Selain itu, dilakukan juga analisis clustering menggunakan metode K-Means dengan jumlah cluster (k) sebanyak dua. Nilai Sum of Squared Errors (SSE) yang diperoleh sebesar 6422,10, dan silhouette score sebesar 0,17. Nilai silhouette yang relatif rendah mengindikasikan bahwa pemisahan antar cluster belum optimal, yang berarti data antar cluster masih tumpang tindih. Hal ini wajar karena dataset ini pada dasarnya lebih sesuai digunakan untuk analisis klasifikasi terawasi dibandingkan untuk clustering tanpa label.