

## PEMBAHASAN UTS DATA MINING TAHUN 2022/2023

Diketahui data sebagai berikut :

Heart Failure Prediction			
Age	Sex	RestingECG	MaxHR
73	F	ST	121
47	M	Normal	118
65	M	ST	122
	M	Normal	157
50	M	ST	156
60	M	ST	99
50	M	Normal	120
43	M	Normal	145

1. Business understanding - Problem apa yang diselesaikan

Penyakit kardiovaskular (CVDs) adalah penyebab kematian nomor 1 secara global, mengambil sekitar 17,9 juta jiwa setiap tahun, yang menyumbang 31% dari semua kematian di seluruh dunia. Empat dari 5 CVD kematian disebabkan oleh serangan jantung dan stroke, dan sepertiga dari kematian ini terjadi sebelum waktunya pada orang di bawah usia 70 tahun. Gagal jantung adalah kejadian umum yang disebabkan oleh CVD dan kumpulan data ini berisi 11 fitur yang dapat digunakan untuk memprediksi kemungkinan penyakit jantung.

Orang dengan penyakit kardiovaskular atau yang berada pada risiko kardiovaskular tinggi (karena adanya satu atau lebih faktor risiko seperti hipertensi, diabetes, hiperlipidemia atau penyakit yang sudah ada) memerlukan deteksi dan manajemen dini di mana model pembelajaran mesin dapat sangat membantu.

2. Data understanding - data berupa jenis apa saja, Attribute apa saja, dan tipe-tipe masing2 Attribute (minimal menggunakan 2 tipe Attribute/data)

Age : usia individu saat dilakukan tes

Sex : jenis kelamin individu (M/F)

ResticECG : restig electrodiogram result

MaxHR : maximum heart rate achieved

3. Data Preparation - Jelaskan 3 contoh penanganan data (missing values, inconsistencies, outliers), dan persiapan data untuk masuk ke model (dicontohkan dengan tabel)

Pada data set ini mengambil 8 entri, namun terlihat beberapa kekosongan pada atribut ph, dan conductivity. Berikut data preparationnya:

### Missing Value

Age
73
47
65
50
60
50
43

Rata-rata:

55,4285714

Age
73
47
65
55
50
60
50
43

setelah di missing value

Outliers : Pada data ini tidak terdapat outliers sehingga data ini aman dan dapat diproses

Incosistencies : Data ini tidak ada inconsistencies data

4. Model - Jenis model apa yang menurut anda bisa digunakan?

Data set ini dimodelkan dengan decision tree untuk menentukan output yaitu penyakit jantung (1) atau tidak (0)