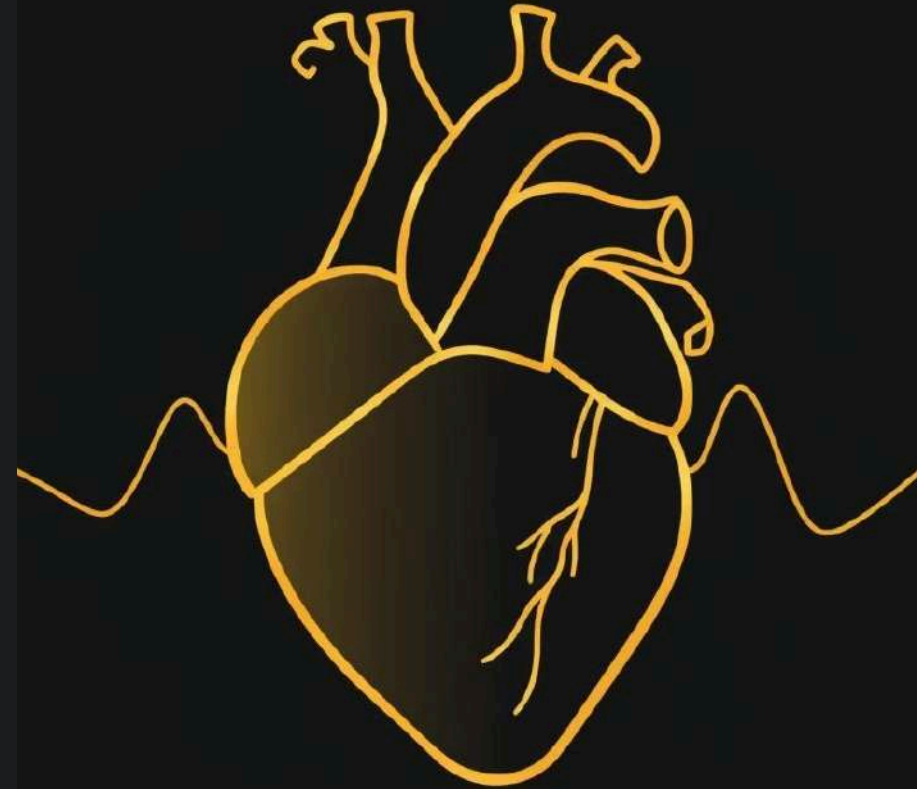


# Klasterisasi Fenotipe Pasien Penyakit Jantung Menggunakan Gaussian Mixture Model (GMM)

**Proyek Analisis Data Kesehatan**

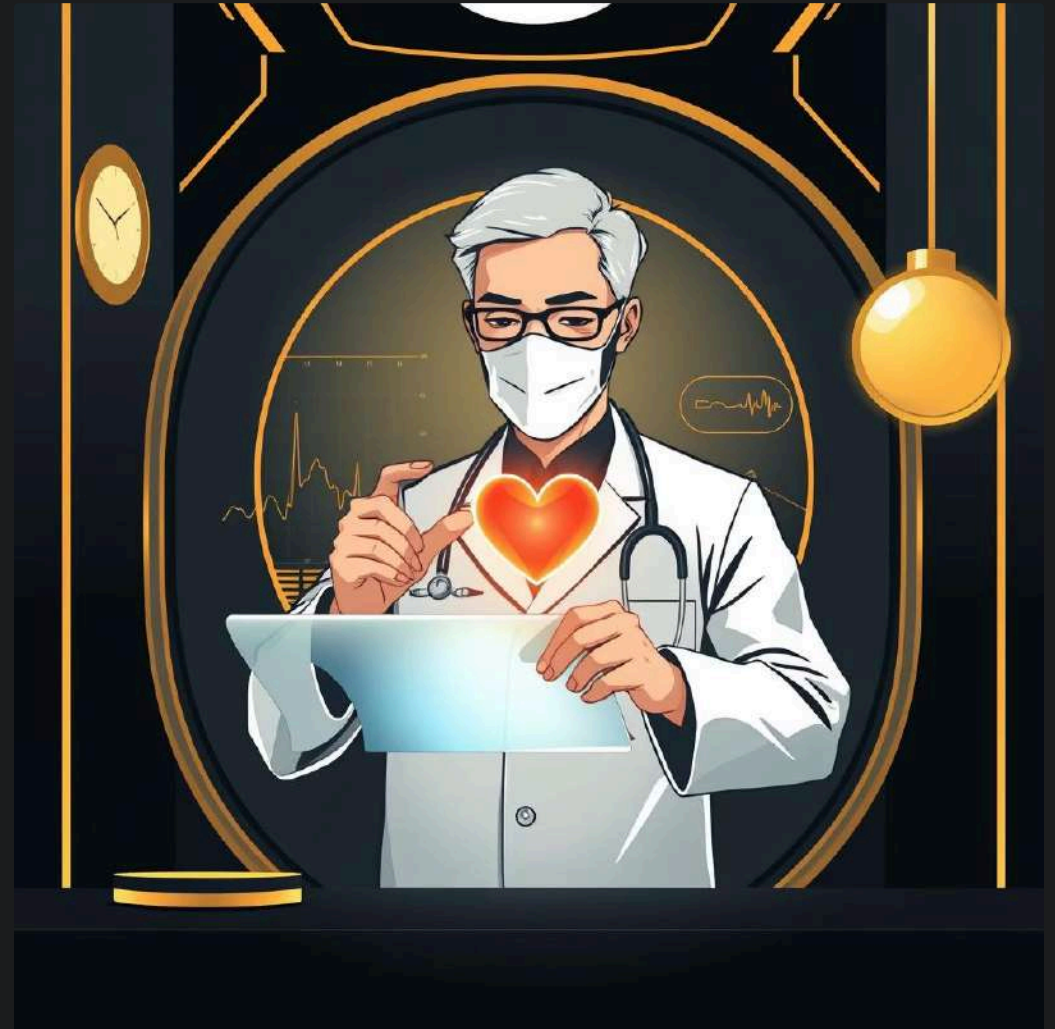
Oleh: Aan Krisnawati dan Gadis Nirmala Sari



# Urgensi Pemetaan Risiko

Penyakit jantung tetap menjadi **penyebab kematian utama** secara global.

- Variabilitas klinis yang sangat tinggi antar pasien.
- Kebutuhan akan personalisasi protokol pengobatan.
- Data multidimensi: usia, tekanan darah, hingga kolesterol.



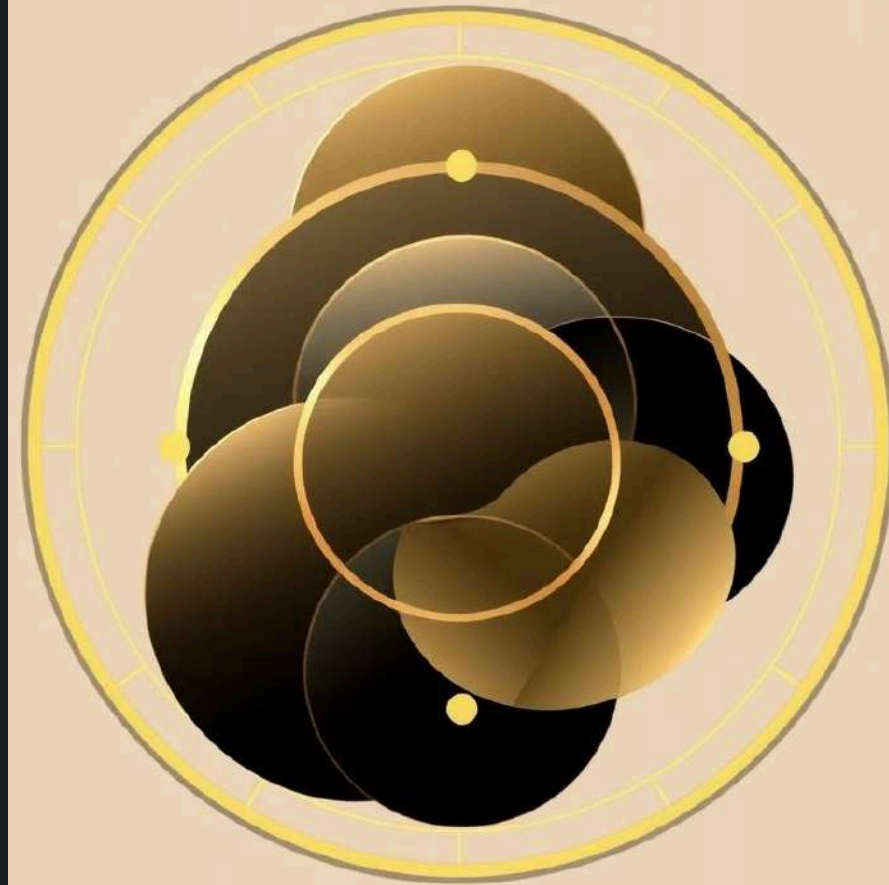
# Rumusan Masalah & Tujuan

## Masalah

Bagaimana mengelompokkan pasien secara otomatis dalam kondisi data yang saling overlap?

## Tujuan Utama

Menerapkan **GMM** untuk identifikasi fenotipe pasien dan interpretasi profil klinis.



# Dataset & Karakteristik Klinis

Menggunakan dataset jantung dengan **918 observasi** dan fitur multidimensi.



## Vital Signs

Age, RestingBP, MaxHR



## Biomarkers

Cholesterol, FastingBS



## Clinical Signs

ChestPainType, ST Slope, Oldpeak

# Pipeline Pra-proses Data



Isi Nilai &  
Outlier

One-Hot  
Encoding

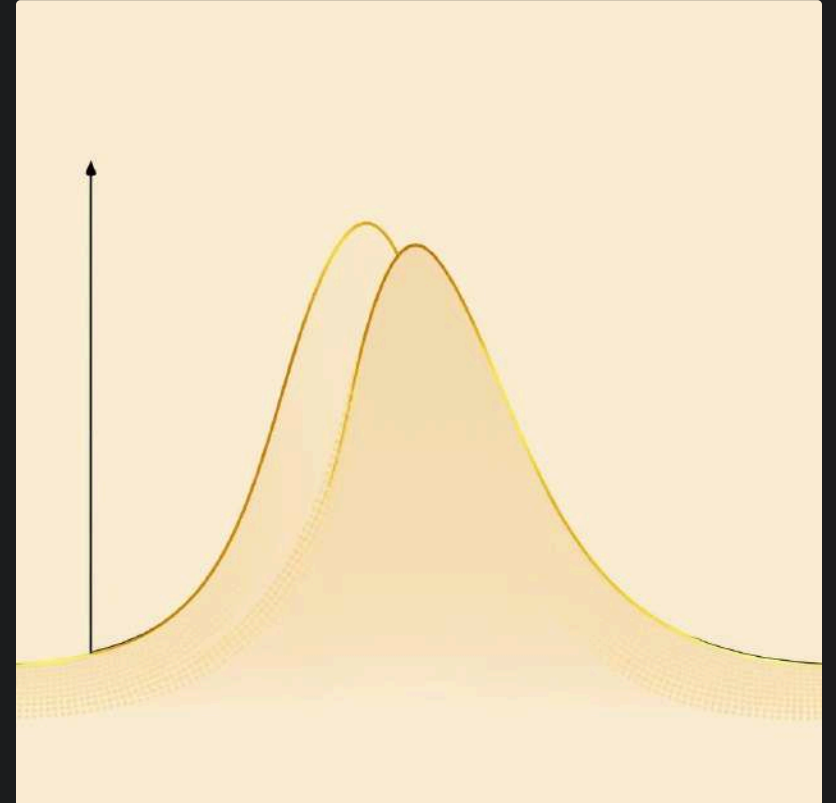
Standardisas  
i Z

❏ Standarisasi sangat krusial bagi GMM agar fitur dengan skala besar tidak mendominasi distribusi Gaussian.

# Mengapa Gaussian Mixture Model?

## Soft Clustering & Fleksibilitas

- **Probabilistik:** Pasien memiliki bobot keanggotaan di setiap kluster.
- **Covariance:** Menangkap bentuk kluster elips, bukan hanya bulat seperti K-Means.
- **Overlap:** Cocok untuk data medis di mana batas kondisi sehat-sakit sering kali kabur.





# Optimasi: Mencari Jumlah Kluster Terbaik

Evaluasi dilakukan pada rentang **k = 1 hingga 10** menggunakan metrik objektif.

## BIC Score

Mencari nilai terendah untuk keseimbangan antara akurasi dan kompleksitas model.



## Silhouette Score

Mengukur pemisahan antar kluster agar tidak terjadi ambiguitas yang tinggi.

# Profil Klaster (Hasil Analisis)

1

## Klaster Risiko Rendah

Usia muda, kolesterol normal, ST slope naik (up-sloping).  
Profil pasien stabil.

2

## Klaster Risiko Tinggi

Usia lanjut, hipertensi, exercise angina (+), dan segmen ST abnormal. Membutuhkan atensi medis segera.

**Output tambahan:** Setiap pasien kini memiliki probabilitas keanggotaan (misal: 85% Klaster 2).



# Insight Klinis & Manfaat



## Segmentasi Pasien

Identifikasi kelompok "borderline" untuk pemantauan preventif.



## Prioritas Intervensi

Alokasi sumber daya rumah sakit yang lebih efektif bagi pasien berisiko tinggi.



## Keputusan Data-Driven

Mendukung dokter dengan bukti kuantitatif dari pola data historis.



# Kesimpulan & Saran

## Kesimpulan

- GMM efektif memetakan fenotipe kompleks pasien jantung.
- BIC & Silhouette memastikan model yang robust dan tidak *overfit*.

## Saran

- Integrasi data gaya hidup & genetika.
- Komparasi dengan metode *Hierarchical Clustering*.

## Terima Kasih