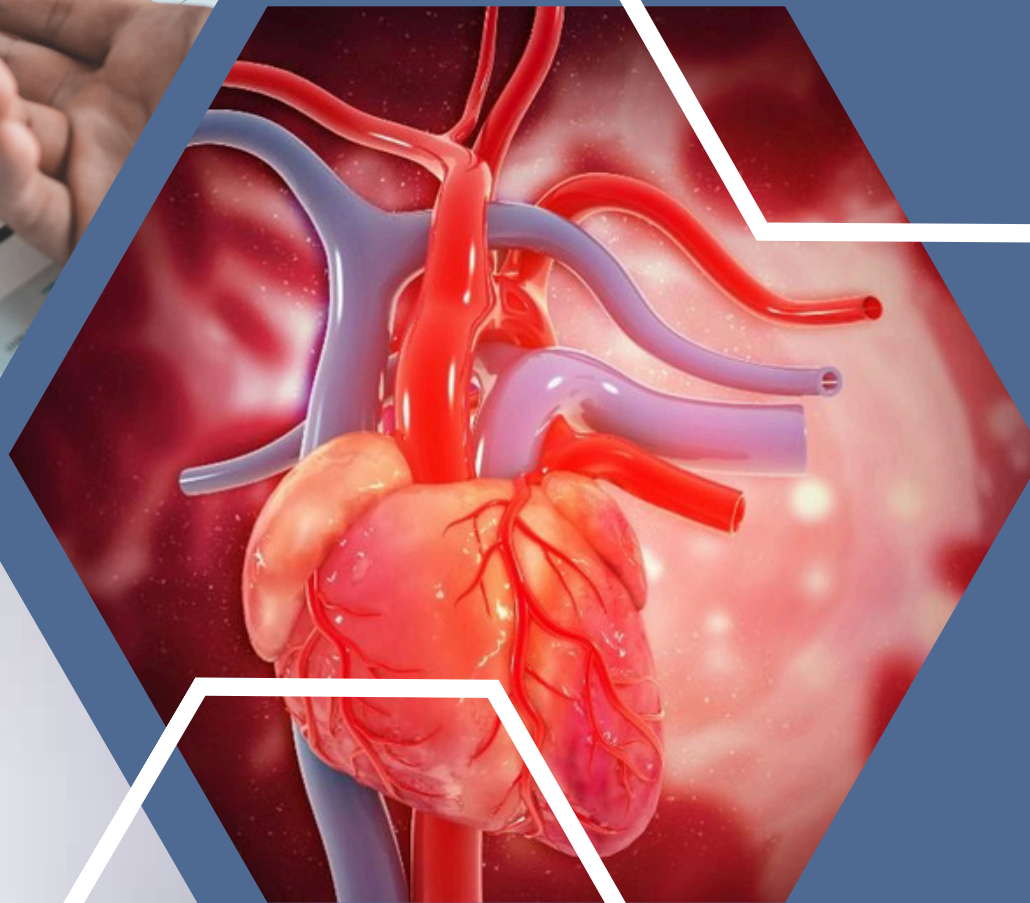


TUGAS KELOMPOK DATA MINING

PREDIKSI GAGAL JANTUNG

Anggota Kelompok :

1. Jasmine Lintang Aurelia (4111422013)
2. Dwi Lufinatul Alifah (4111422014)
3. Ivana Qotrunnada Shintia Maramis (4111422025)
4. Yolanda Agnes Halbet Sitompul (4111422027)



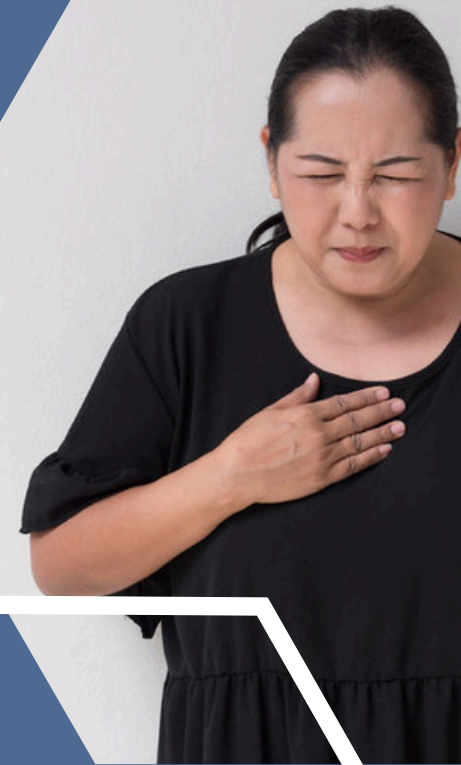
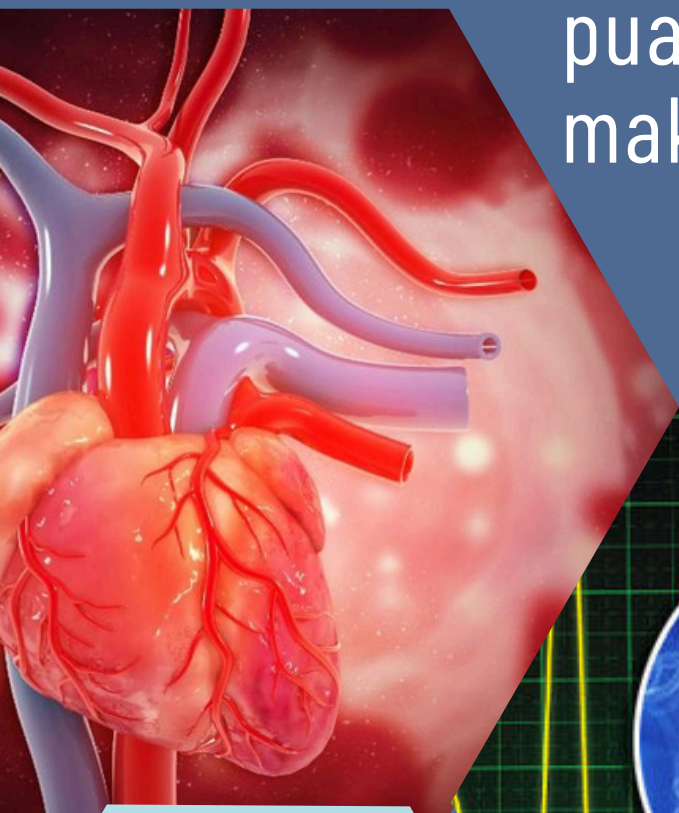
BUSINESS UNDERSTANDING

Penyakit kardiovaskular (CVD) merupakan masalah kesehatan utama di dunia dan penyebab utama kematian dengan sekitar 17,9 juta jiwa meninggal setiap tahun, yang setara dengan 31% dari total kematian global. Gagal jantung adalah salah satu akibat umum dari CVD. Orang dengan penyakit kardiovaskular atau risiko tinggi (seperti hipertensi, diabetes, atau hiperlipidemia) memerlukan deteksi dan manajemen dini. Model machine learning dapat membantu mengatasi masalah ini secara otomatis.

Tujuannya untuk mengklasifikasikan/memprediksi apakah seorang pasien rentan terhadap gagal jantung berdasarkan beberapa faktor serta klasifikasi biner dengan beberapa fitur numerik dan kategorikal.

DATA UNDERSTANDING

Dari data yang kami ambil di kaggle tentang prediksi penyakit gagal jantung, ada beberapa faktor yang memengaruhi gagal jantung pada seseorang serta parameter kesehatannya, seperti usia, jenis kelamin, jenis nyeri dada, tekanan darah, kolesterol, gula darah puasa, hasil elektrokardiogram, serta detak jantung maksimum yang dicapai.



DATA PREPARATION

Berdasarkan data yang kami punya, kami telah melakukan preprocessing dengan mencari missing value, one hot encoding dan visualisasi data.

MODELLING

Modelling yang kita lakukan yaitu menerapkan clustering K-Means dengan memasangkan :

1. Cholesterol (penyakit kolesterol) dengan Age (usia)
2. RestingBP (tekanan darah istirahat) dengan Age (usia)
3. MaxHR (detak jantung maksimal) dengan RestingBP (tekanan darah istirahat).

Pada setiap clustering K-Means dibagi menjadi 3 cluster yang tujuannya adalah untuk melihat pengaruh dari beberapa faktor yang ada terhadap penyakit jantung.



EVALUATION

Dari modelling yang telah kami lakukan dengan menggunakan clustering K-Means. Hasil clustering tersebut untuk mengklasifikasikan/memprediksi apakah seorang pasien rentan terhadap gagal jantung berdasarkan beberapa faktor yang telah kami lakukan. Kami melakukan evaluasi dengan menggunakan logistic regression yang menghasilkan akurasi mendekati 85%.



The background features a collage of medical images related to heart health, including a glowing blue heart with an ECG line, a person clutching their chest in pain, and a detailed anatomical heart model. These images are framed by a pattern of overlapping hexagons in various shades of blue and teal.

SUMMARY

Dari Modelling dan preprocessing yang telah kami lakukan pada data tersebut saran yang dapat kami berikan adalah pada data faktor pengaruh penyakit gagal jantung bisa dilakukan penelitian lebih lanjut dengan melihat faktor-faktor lain yang berpengaruh, misalnya berat badan, faktor keturunan, dan faktor pola hidup.

Karena kami telah melakukan Clustering K-Means hanya melihat pada beberapa faktor. Saran yang dapat diberikan pada Modelling penggunaan model bisa dilakukan dengan menggunakan Clustering K-Means berdasarkan semua faktor yang ada sehingga bisa melihat lebih jelas pengaruh dari semua faktor terhadap gagal jantung.