

ANALISIS PREDIKTIF PENYAKIT JANTUNG

MENGGUNAKAN LIMA MODEL MACHINE

LEARNING DALAM MENENTUKAN PELUANG

SESEORANG TERKENA PENYAKIT JANTUNG

2602079161 - Felicia Angel Wijaya

2602093683 - Lambert Aditama Soehardjianto

2602207054 - Thaddeus Kendrick Andrian

2602079041 - Wilsen Soetresno

LATAR BELAKANG

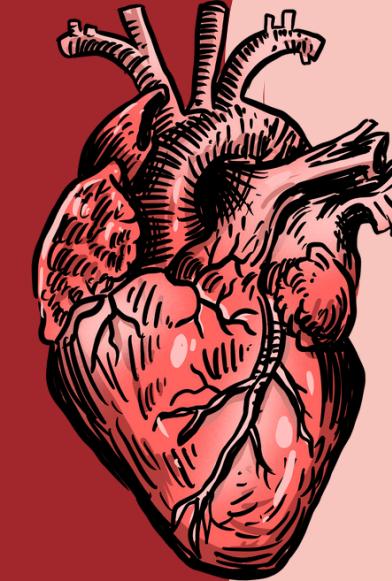
Di zaman modern ini, penyakit jantung merupakan penyakit yang menyebabkan kematian terbesar di dunia. Menurut Organisasi Kesehatan Dunia (WHO), penyakit kardiovaskular, menyebabkan lebih dari 17,9 Juta kematian setiap tahunnya. Penyakit jantung dapat dipicu oleh masalah internal seperti memiliki riwayat penyakit lainnya yang men-trigger penyakit jantung, serta kondisi eksternal seperti gaya hidup yang tidak sehat.

Pada studi kali ini, kami hendak menggunakan dataset pasien yang akan dimanfaatkan dalam memvisualisasi data untuk melihat faktor-faktor yang menyebabkan seseorang terkena penyakit jantung, serta membuat prediksi terkait kemungkinan seseorang terkena penyakit jantung dengan pemanfaatan lima model *Machine Learning*. Dari kelima model tersebut, kami akan menganalisis hasilnya dan mencari tahu model manakah yang paling akurat dalam memprediksi peluang terkena penyakit.



RUMUSAN MASALAH

- Apa faktor-faktor yang dapat menyebabkan seseorang terserang penyakit jantung?
- Bagaimana cara memanfaatkan model *Machine Learning* untuk memprediksi risiko penyakit jantung berdasarkan berbagai faktor risiko yang dipertimbangkan?
- Apa model *Machine Learning* yang paling akurat dalam penelitian prediktif resiko terkena penyakit jantung?



TUJUAN PENELITIAN

- Mencari tahu faktor-faktor yang dapat menyebabkan seseorang terserang penyakit jantung.
- Mengetahui cara memanfaatkan model *Machine Learning* untuk memprediksi risiko penyakit jantung berdasarkan berbagai faktor risiko yang dipertimbangkan.
- Mengetahui model *Machine Learning* yang paling akurat dalam penelitian prediktif resiko terkena penyakit jantung.



MANFAAT PENELITIAN

mengetahui faktor-faktor
yang memicu seseorang
terkena penyakit jantung

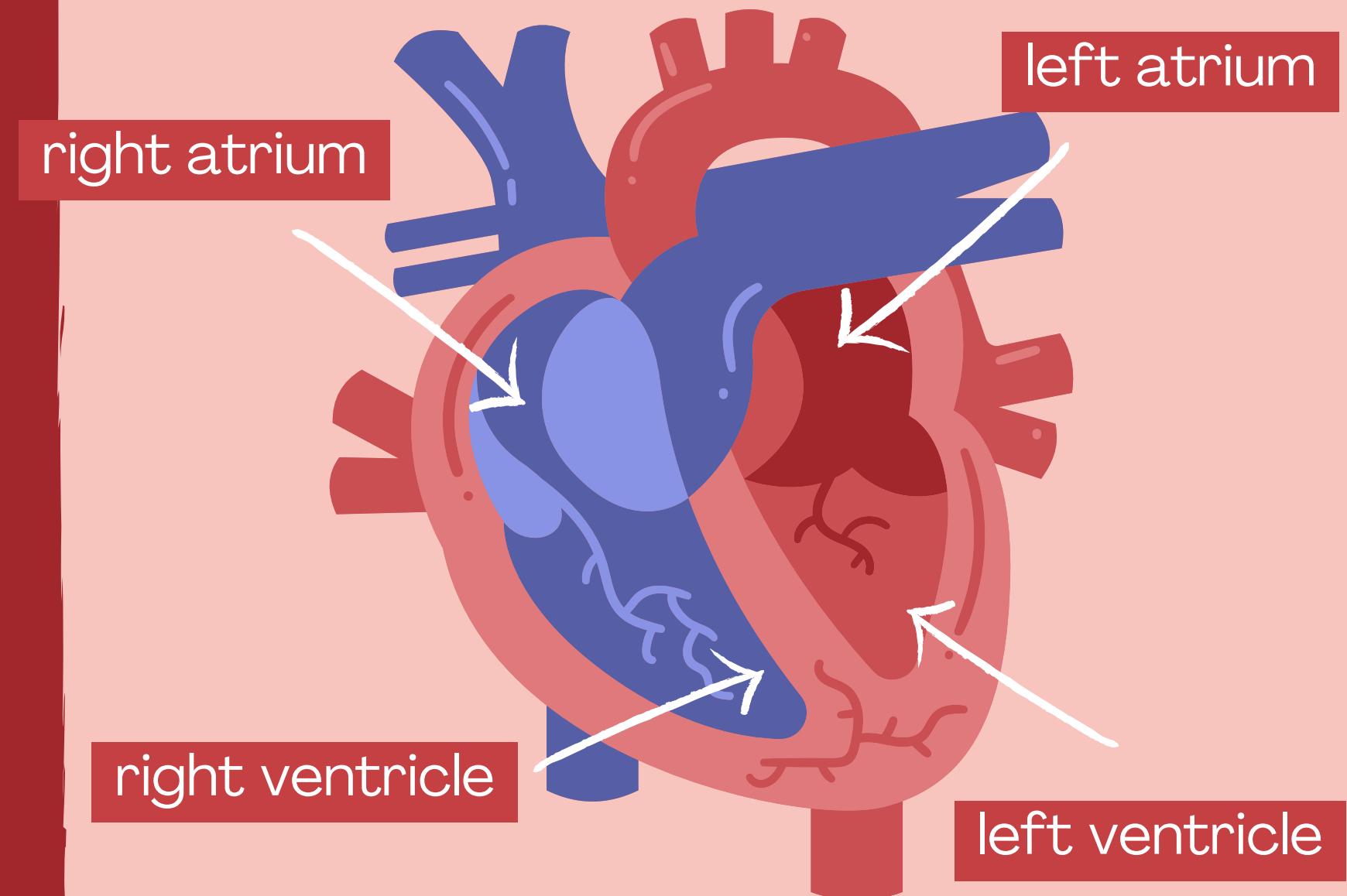
menunjukkan penggunaan
Machine Learning dalam analisa
dan penelitian prediktif

menciptakan kolaborasi
antara bidang medis dengan
bidang teknologi

meningkatkan kewaspadaan
pada masyarakat umum untuk
menjaga kesehatan jantung

THE HEART

Jantung adalah organ otot berongga yang terletak di tengah dada dan berfungsi sebagai pompa untuk memompa darah ke seluruh tubuh. Fungsi utama jantung adalah memompa darah yang kaya oksigen dari paru-paru ke seluruh tubuh melalui sistem peredaran darah.



HEART DISEASE



GEJALA

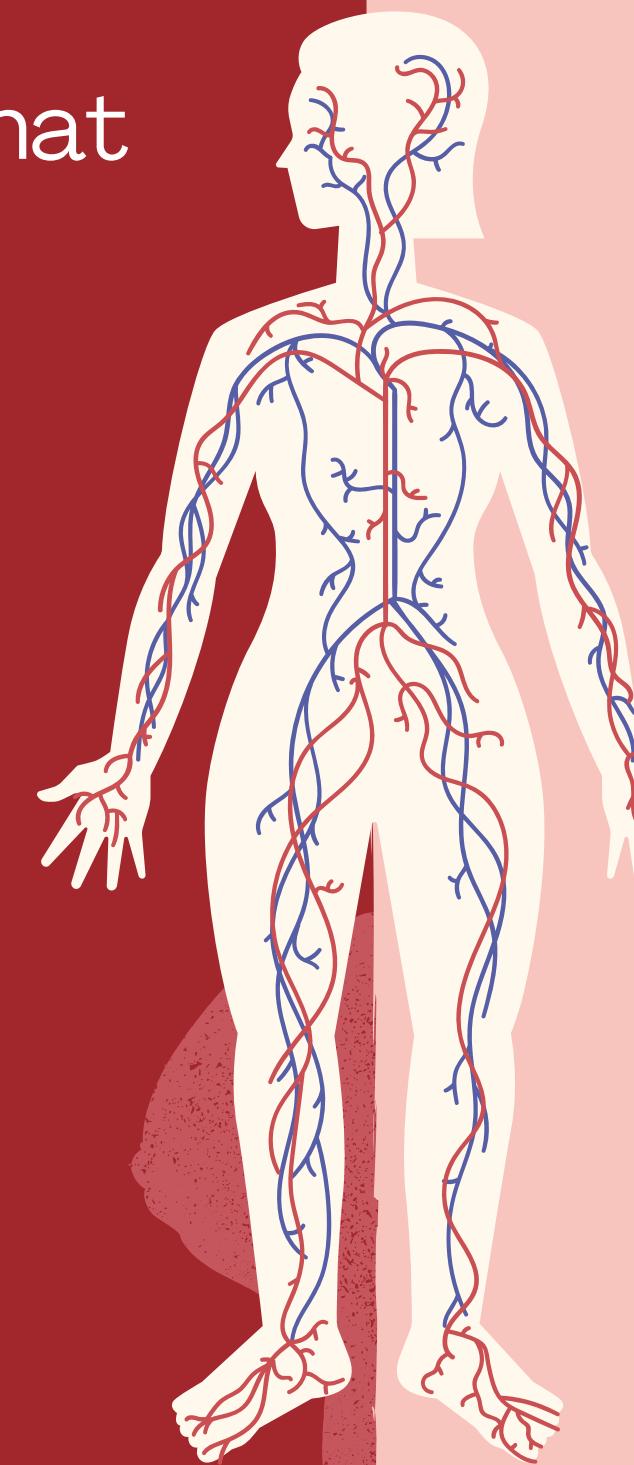
- nyeri dada (angina)
- sesak nafas
- mudah lelah
- palpitasi (detak jantung yang tidak teratur)
- edema (pembengkakan)
- gagal jantung

Penyakit jantung adalah kondisi medis yang melibatkan gangguan pada jantung atau pembuluh darah yang memasok darah ke jantung, yang terjadi ketika adanya penumpukan plak aterosklerosis (aterosklerosis), gangguan irama jantung (aritmia), kelainan katup jantung, atau kelainan bawaan pada struktur jantung. Penyakit jantung dapat terjadi pada siapa saja, termasuk pria, wanita, dan bahkan anak-anak.



TERKENDALI

- 1 Gaya hidup yang tidak sehat
- 2 Diabetes
- 3 Hipertensi
- 4 Obesitas

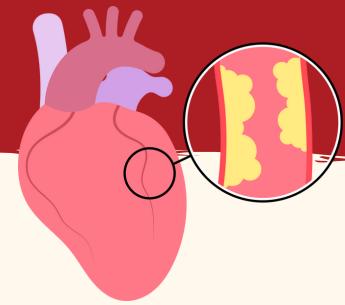


FAKTOR

TIDAK TERKENDALI

- 1 Usia
- 2 Jenis kelamin
- 3 Genetik (Keturunan)

JENIS PENYAKIT JANTUNG



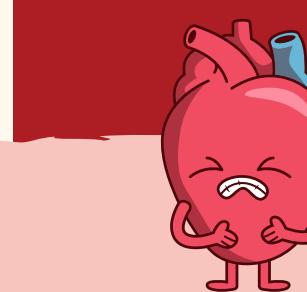
Jantung Konorer

Arteri koroner yang memasok darah ke jantung mengalami penyempitan oleh plak atherosklerosis, membuat pasokan darah dan oksigen ke jantung menjadi terbatas.



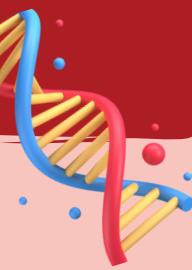
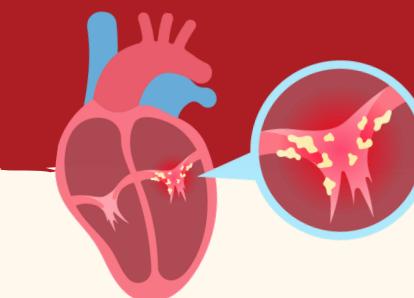
Gagal Jantung

Ketidakmampuan jantung untuk dapat memompa darah dengan cukup kuat untuk memenuhi kebutuhan tubuh.



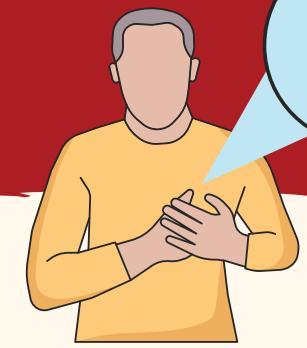
Penyakit Katup Jantung

Salah satu atau beberapa katup jantung mengalami kelainan, seperti penyempitan (stenosis) atau kebocoran (regurgitasi).



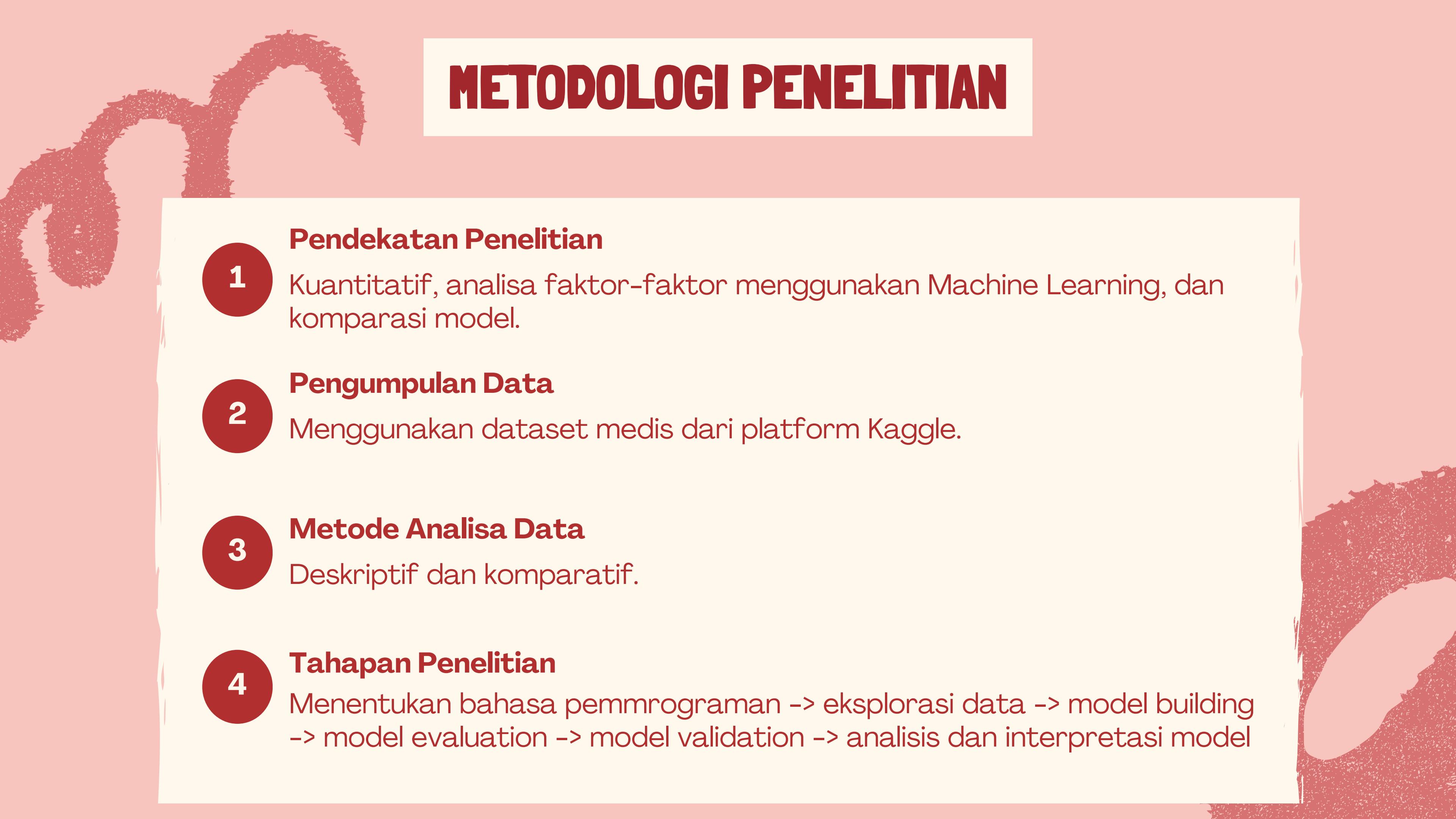
Penyakit Jantung Bawaan

Penyakit jantung bawaan adalah kelainan struktural pada jantung yang hadir sejak lahir. Contoh kelainan meliputi lubang di dinding jantung (septum), kelainan katup, atau susunan abnormal pembuluh darah.



Aritmia

gangguan irama jantung, di mana jantung dapat berdetak terlalu cepat (takikardia), terlalu lambat (bradikardia), atau tidak teratur.



METODOLOGI PENELITIAN

Pendekatan Penelitian

1 Kuantitatif, analisa faktor-faktor menggunakan Machine Learning, dan komparasi model.

Pengumpulan Data

2 Menggunakan dataset medis dari platform Kaggle.

Metode Analisa Data

3 Deskriptif dan komparatif.

Tahapan Penelitian

4 Menentukan bahasa pemrograman -> eksplorasi data -> model building -> model evaluation -> model validation -> analisis dan interpretasi model

SUMBER DATASET



Heart Failure Prediction Dataset

Dataset · 3y ago · by [fedesoriano](#)

Heart failure is a common event caused by CVDs and this dataset contains 11 features that can be used

2529

151,246 downloads

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
2	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
5	39	M	NAP	120	339	0	Normal	170	N	0.0	Up	0

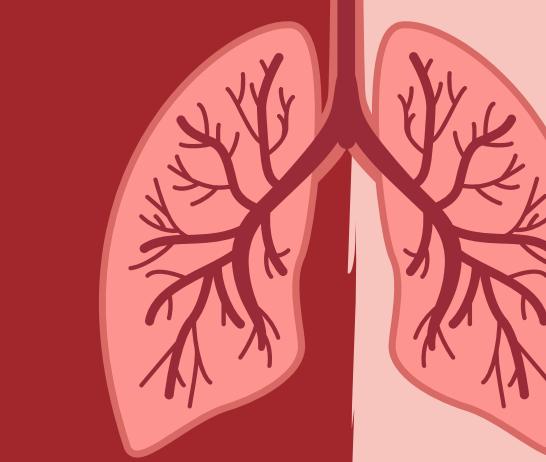
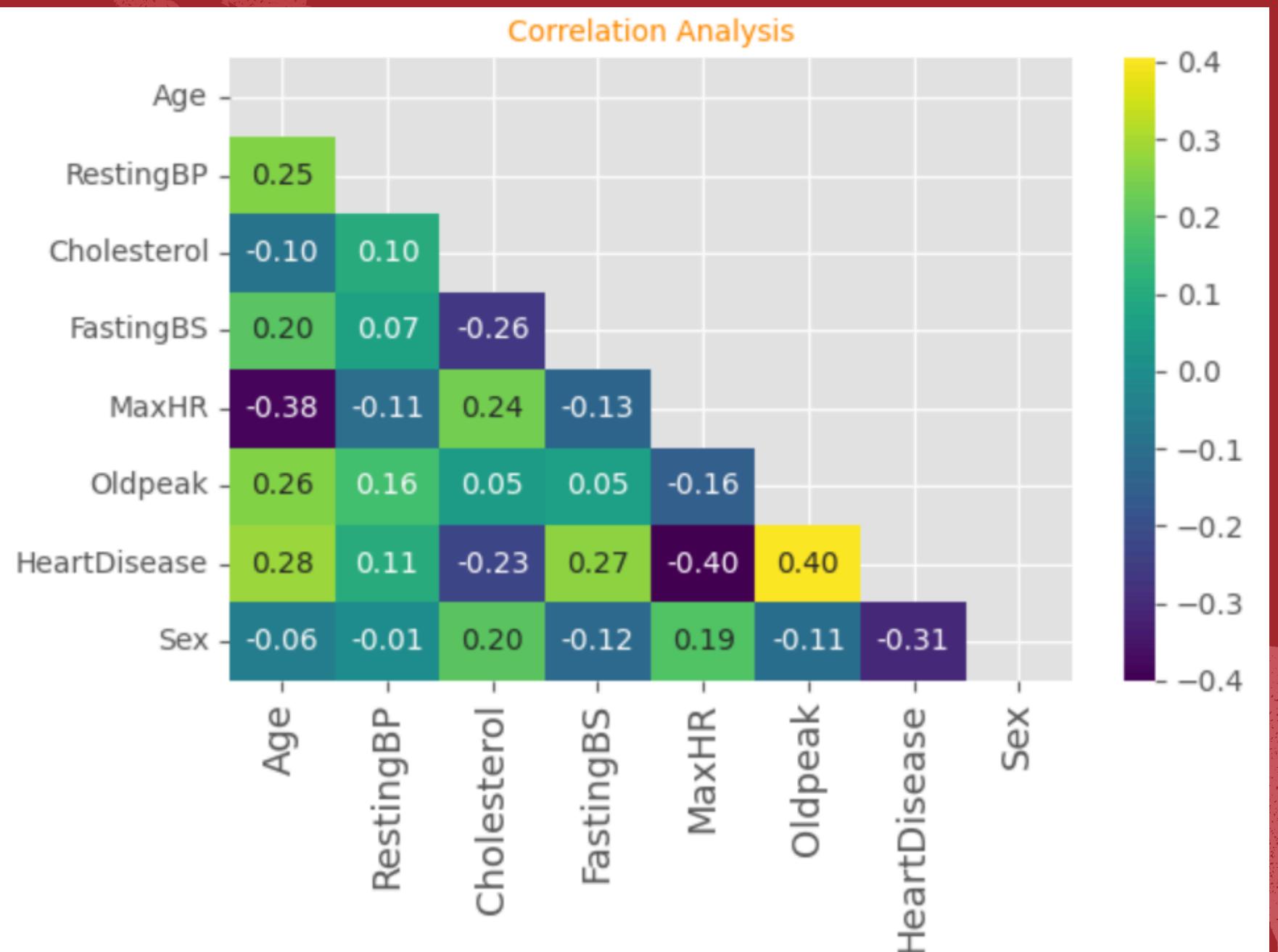
Total Observations

918

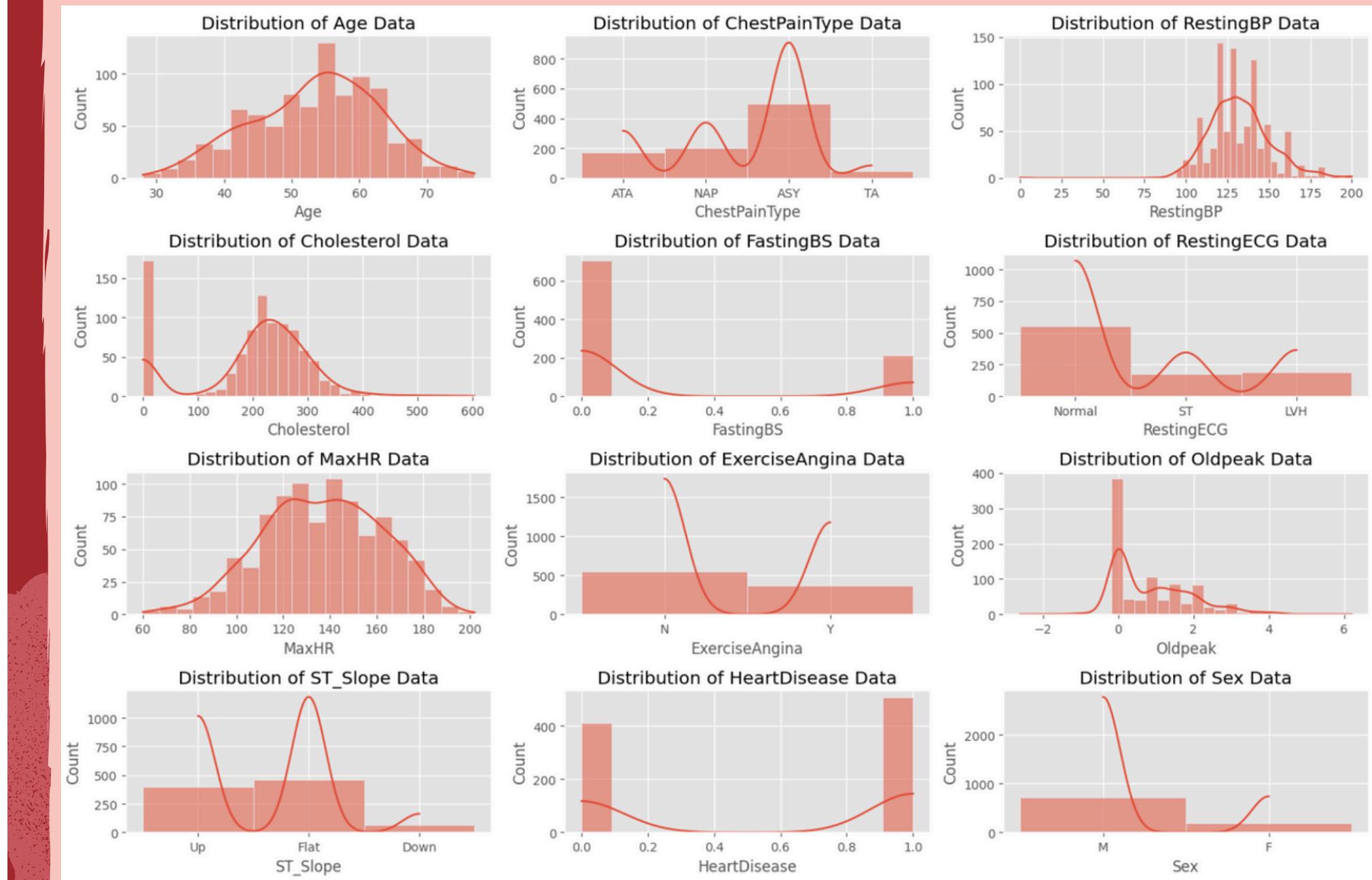
VARIABEL

- 1 **Age:** age of the patient [years]
- 2 **Sex:** sex of the patient [M: Male, F: Female]
- 3 **ChestPainType:** chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
- 4 **RestingBP :** tekanan darah istirahat [mmHg]
- 5 **Cholesterol:** serum cholesterol [mg/dl]
- 6 **FastingBS** (Kadar gula darah puasa): fasting blood sugar [1: if FastingBS > mg/dl, 0: otherwise]
- 7 **RestingECG:** resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]
- 8 **MaxHR:** maximum heart rate achieved [Numeric value between 60 and 202]
- 9 **ExerciseAngina:** exercise-induced angina [Y: Yes, N: No]
- 10 **ST_Slope:** the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]
- 11 **Oldpeak:** ST [Numeric value measured in depression]
- 12 **HeartDisease:** output class [1: heart disease, 0: Normal]

CORRELATION MATRIX



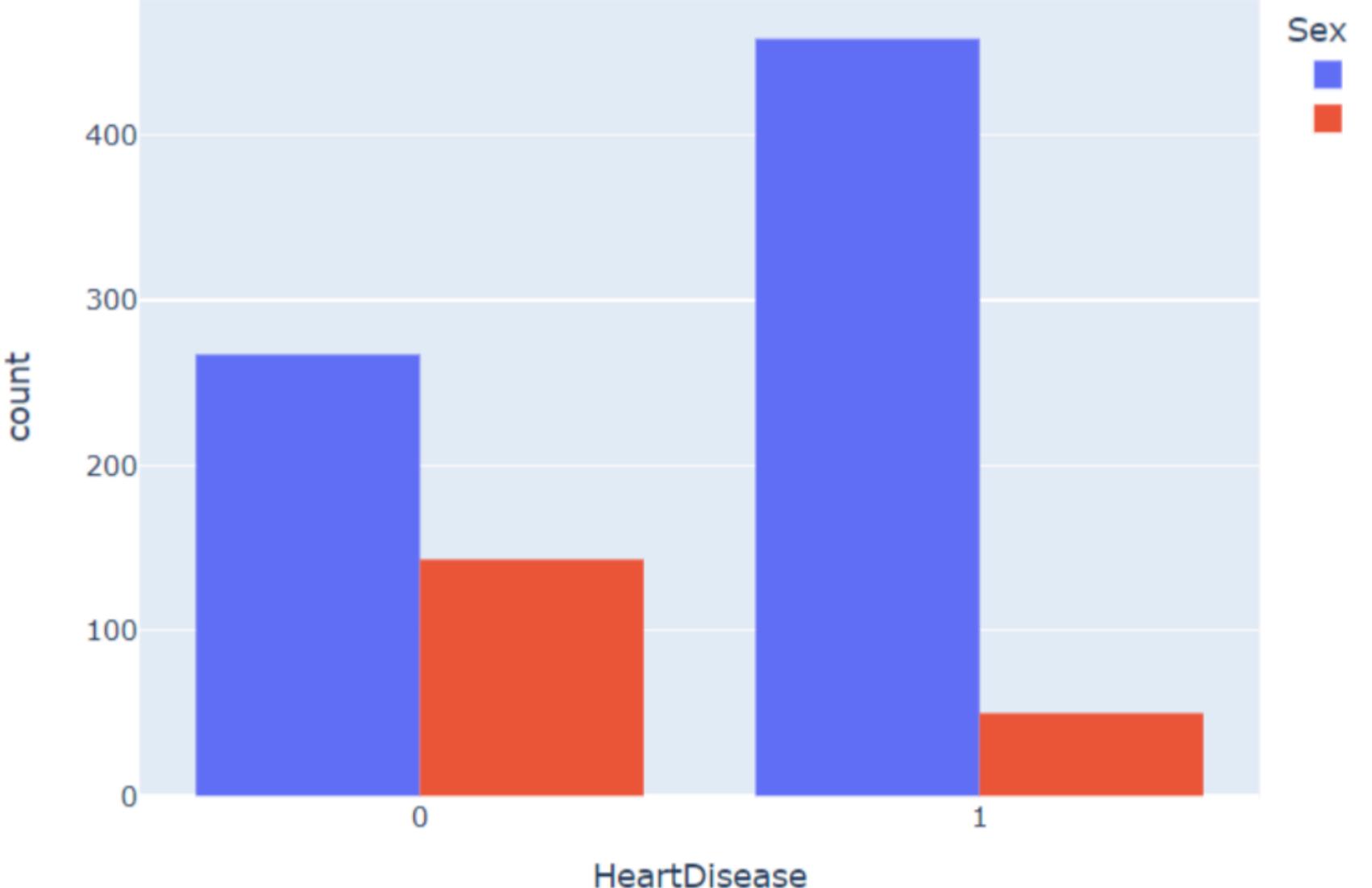
UNIVARIATE DISTRIBUTION



WHAT CAUSES HEART DISEASE ?

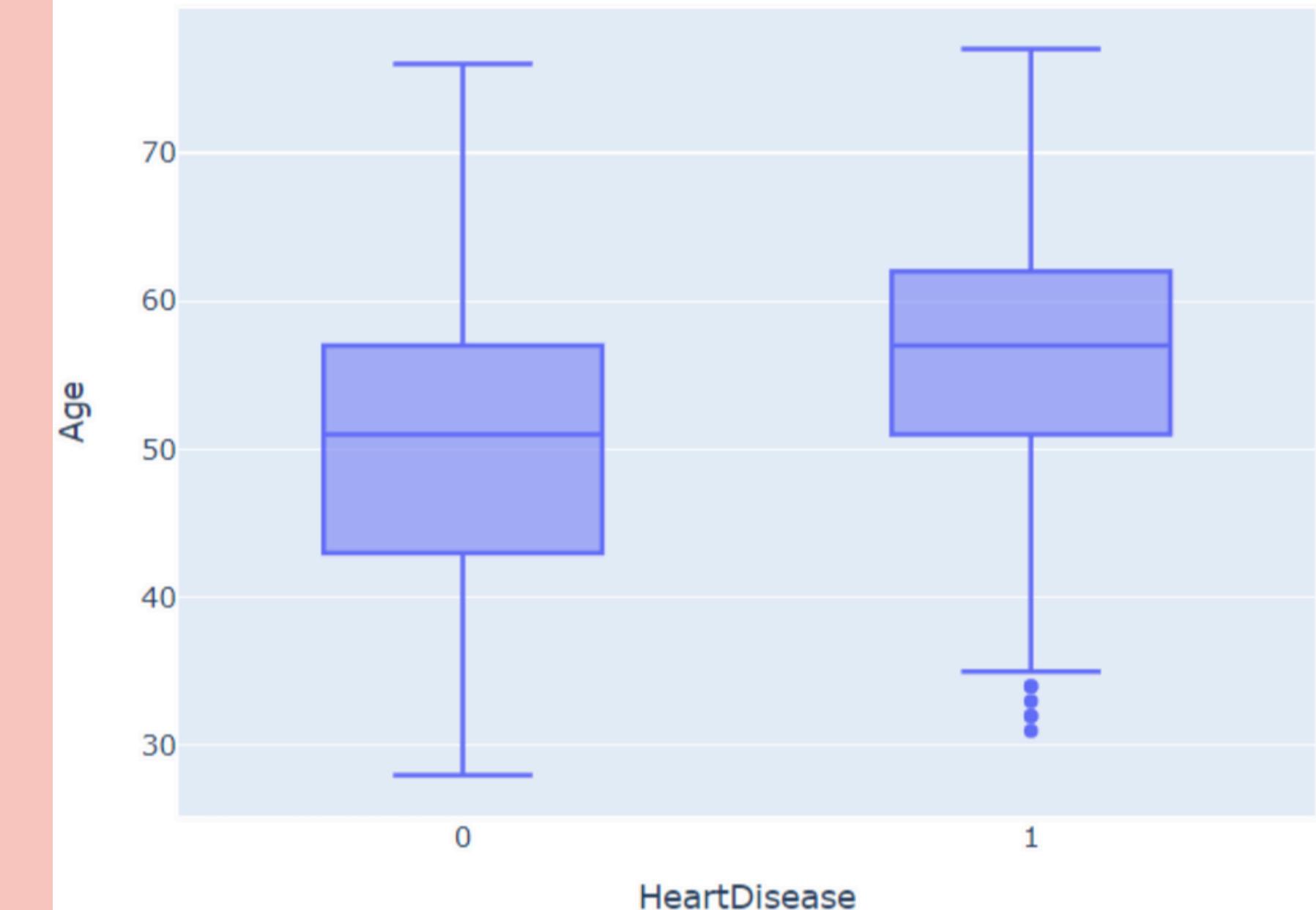
SEX VS HEART DISEASE

Distribution of Heart Diseases

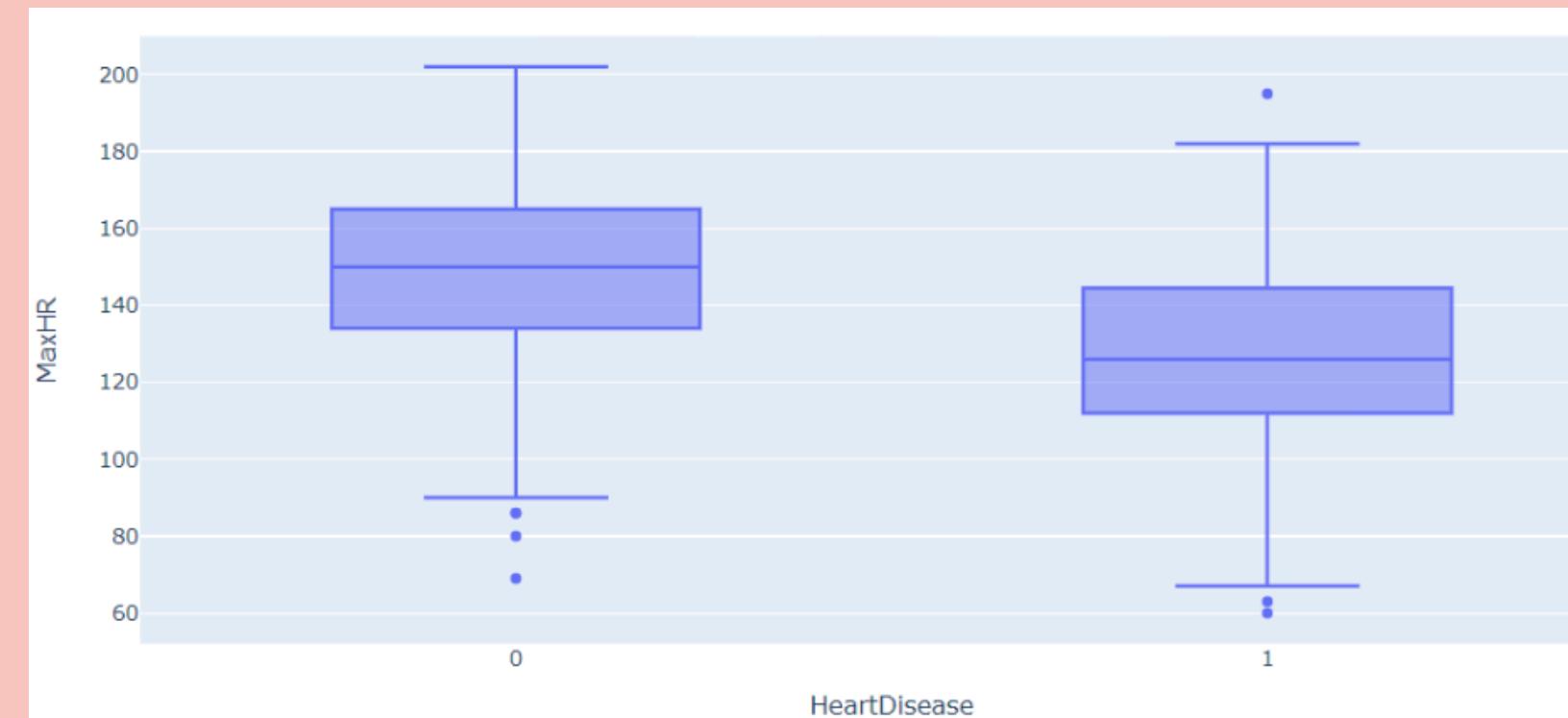


AGE VS HEART DISEASE

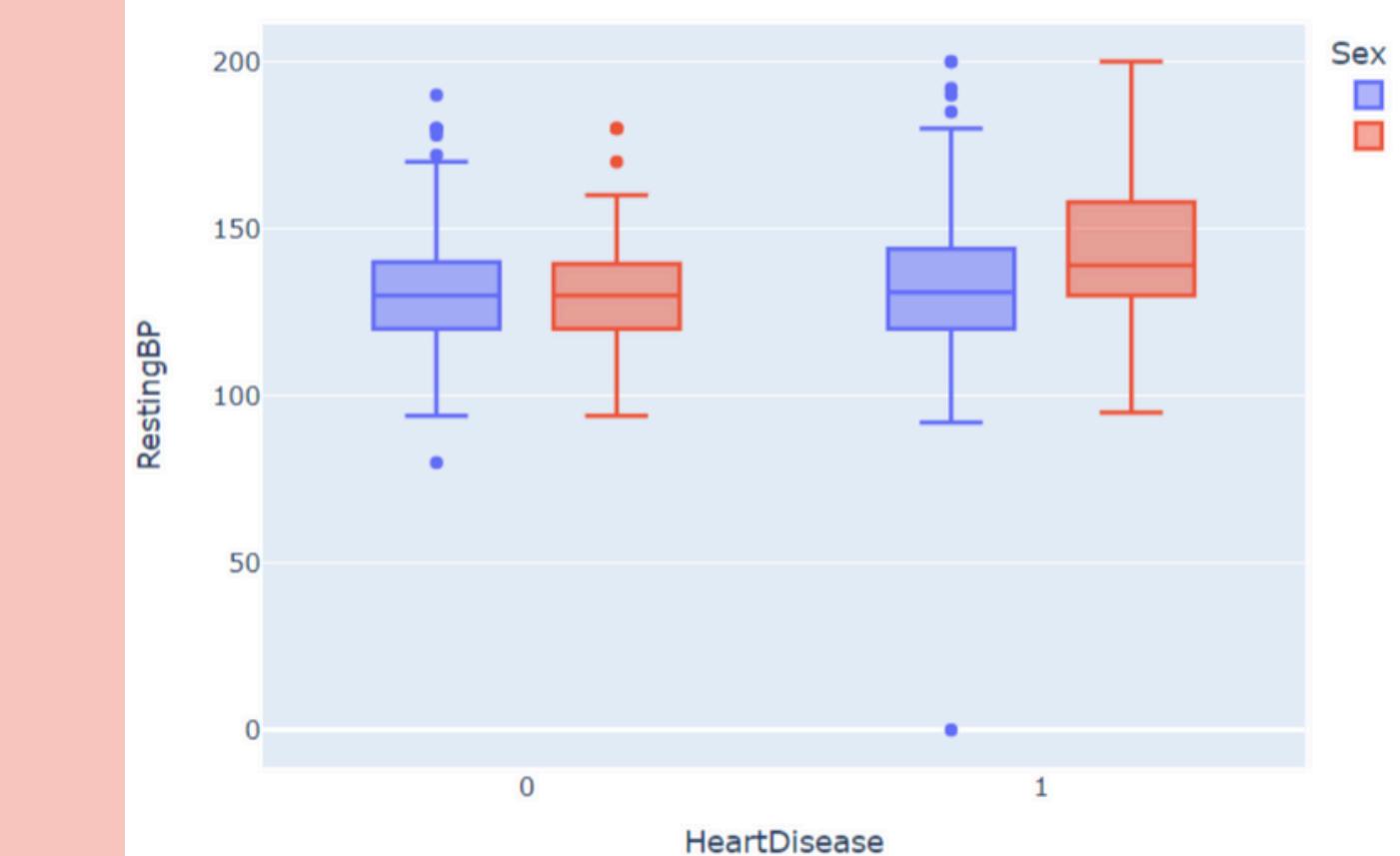
Distrubution of Age



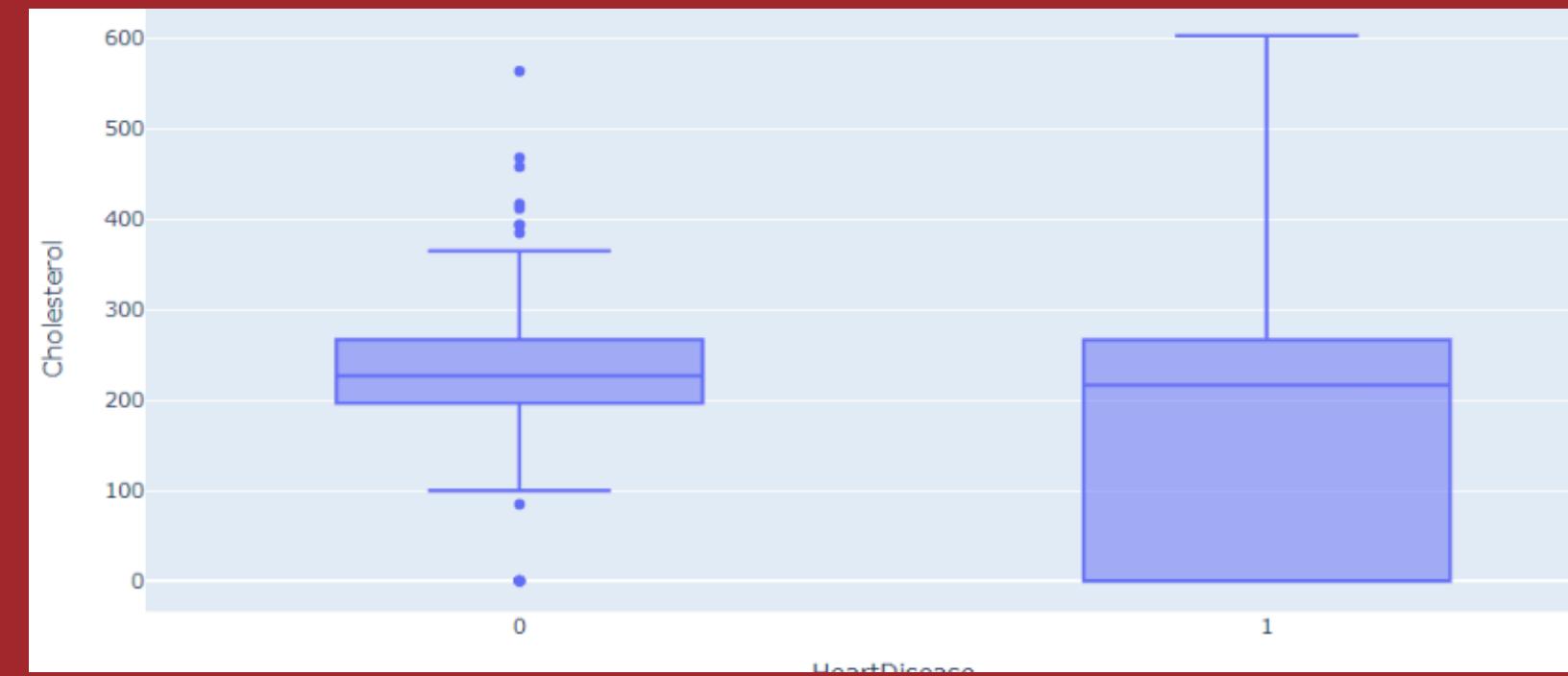
MAXIMUM HEART RATE VS HEART DISEASE



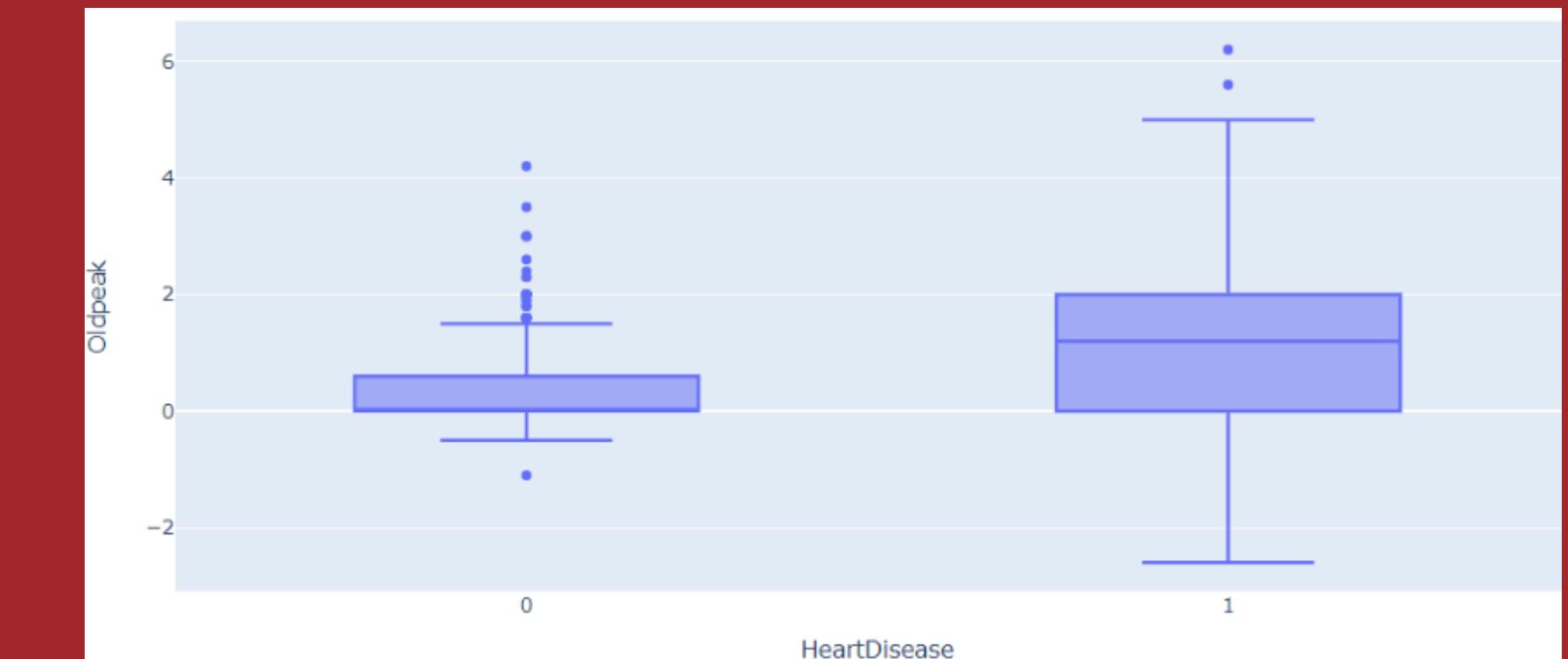
RESTINGBP VS HEART DISEASE



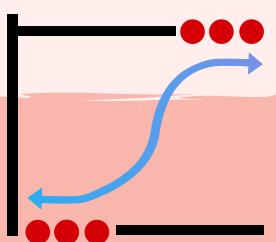
CHOLESTEROL VS HEART DISEASE



OLDPEAK VS HEART DISEASE

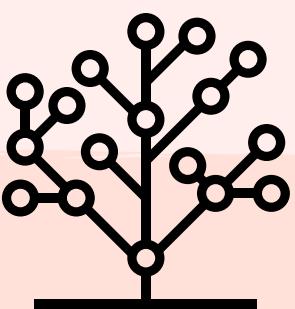


JENIS MACHINE LEARNING YANG DIGUNAKAN



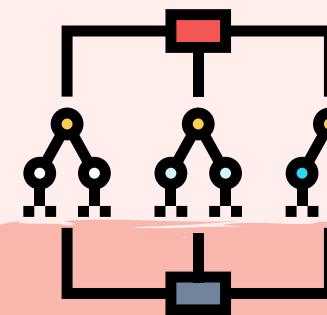
Logistic Regression

Model yang digunakan untuk memprediksi suatu hasil akhir yang berbentuk binary output (1 dan 0, iya dan tidak) berdasarkan kondisi atau atribut yang ada. Hasil output ini dilakukan berdasarkan perhitungan probabilitas hasil akhir berdasarkan korelasi faktor-faktor serta nilai dari faktor-faktor terhadap hasil akhir.



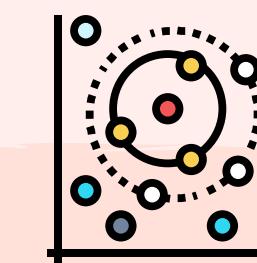
Decision Tree

Model yang digunakan untuk model klasifikasi dan regresi. Decision Tree Algorithm memprediksi nilai output berdasarkan serangkaian kondisi atau atribut yang ada dalam dataset.



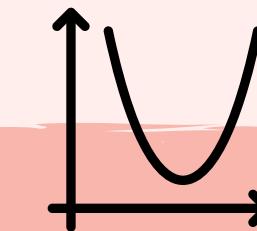
Random Forest

Model yang bekerja dengan cara membangun beberapa pohon keputusan dari subset acak data pelatihan dan fitur. Kemudian, seluruh prediksi dari pohon-pohon akan digabungkan melalui voting untuk meningkatkan kinerja prediksi secara keseluruhan.



K-Nearest Neighbors

Model yang menggunakan kedekatan untuk membuat klasifikasi atau prediksi tentang pengelompokan titik data individual dengan asumsi bahwa titik serupa dapat ditemukan berdekatan satu sama lain. Algoritma ini hanya membutuhkan penyimpanan data pelatihan untuk prediksi & sangat bergantung pada parameter k dan metrik jarak yang tepat.



Stochastic Gradient Descent Model

Model memperbarui parameter model secara iteratif untuk meminimalkan fungsi kerugian. Pada setiap iterasi, SGD menggunakan satu contoh data pelatihan yang dipilih secara acak untuk memperbarui parameter, yang membuatnya lebih efisien dan cepat untuk dataset besar. Proses ini terus dilakukan hingga model mencapai jumlah iterasi yang telah ditentukan tercapai.

1

LOGISTIC REGRESSION

Training Accuracy : 87.33 %
 Model Accuracy Score : 86.96 %
 Confusion Matrix:
 [[63 13]
 [11 97]]

Classification_Report:

	precision	recall	f1-score	support
0	0.85	0.83	0.84	76
1	0.88	0.90	0.89	108
accuracy			0.87	184
macro avg	0.87	0.86	0.86	184
weighted avg	0.87	0.87	0.87	184

Validation Score: 86.36 %
 Std. Dev: 4.37 %

2

DECISION TREE

Training Accuracy : 100.0 %
 Model Accuracy Score : 82.61 %
 Confusion Matrix:
 [[62 14]
 [18 90]]

Classification_Report:

	precision	recall	f1-score	support
0	0.78	0.82	0.79	76
1	0.87	0.83	0.85	108
accuracy			0.83	184
macro avg	0.82	0.82	0.82	184
weighted avg	0.83	0.83	0.83	184

Validation Score: 79.55 %
 Std. Dev: 5.98 %

3

RANDOM FOREST

Training Accuracy : 100.0 %
 Model Accuracy Score : 88.04 %
 Confusion Matrix:
 [[64 12]
 [10 98]]

Classification_Report:

	precision	recall	f1-score	support
0	0.86	0.84	0.85	76
1	0.89	0.91	0.90	108
accuracy			0.88	184
macro avg	0.88	0.87	0.88	184
weighted avg	0.88	0.88	0.88	184

Validation Score: 86.23 %
 Std. Dev: 5.12 %

4

K-NEAREST NEIGHBORS

Training Accuracy : 91.14 %
 Model Accuracy Score : 84.78 %
 Confusion Matrix:
 [[63 13]
 [15 93]]

Classification_Report:

	precision	recall	f1-score	support
0	0.81	0.83	0.82	76
1	0.88	0.86	0.87	108
accuracy			0.85	184
macro avg	0.84	0.85	0.84	184
weighted avg	0.85	0.85	0.85	184

Validation Score: 84.46 %
 Std. Dev: 4.04 %

STOCHASTIC GRADIENT DESCENT (SGD)

Training Accuracy : 81.47 %

Model Accuracy Score : 80.98 %

Confusion Matrix:

```
[[52 24]
 [11 97]]
```

Classification_Report:

	precision	recall	f1-score	support
0	0.83	0.68	0.75	76
1	0.80	0.90	0.85	108
accuracy			0.81	184
macro avg	0.81	0.79	0.80	184
weighted avg	0.81	0.81	0.81	184

Validation Score: 83.93 %

Std. Dev: 3.90 %

MODEL COMPARISON

Model	Training Accuracy	Model Accuracy Score	Model f1 Score
Random Forest	100.00	88.04	89.91
Logistic Regression	87.33	86.96	88.99
Decision Tree	100.00	82.61	84.91
K-Nearest Neighbors	91.14	84.78	84.78
Stochastic Gradient Descent	81.47	80.98	80.98

Tabel diatas menunjukan keefektifan masing-masing *Machine Learning Algorithm* yang diurutkan berdasarkan F1-score.

Dapat dilihat bahwa semua algoritma yang kami gunakan memiliki tingkat akurasi yang tinggi, dengan minimum 81, dan maksimum 90.

Machine Learning Algorithm yang paling akurat untuk kasus ini adalah Random Forest.

KESIMPULAN

- Faktor-Faktor yang paling berpengaruh pada penyakit jantung adalah :
 - Usia
 - Sex
 - Gula darah
 - Oldpeak (ST Depression)
- *Machine Learning Algorithm* yang paling akurat untuk kasus ini adalah *random forest algorithm*.



**z
tra
9
rak**

