

**EVALUASI KINERJA METODE *SUPPORT VECTOR MACHINE* (SVM),
NAIVE BAYES DAN *DECISION TREE* UNTUK DIAGNOSA PENYAKIT
JANTUNG**

(Skripsi)

Oleh

SENDY HANI PRAMITA



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

**EVALUASI KINERJA METODE *SUPPORT VECTOR MACHINE* (SVM),
NAIVE BAYES DAN *DECISION TREE* UNTUK DIAGNOSA PENYAKIT
JANTUNG**

Oleh

SENDY HANI PRAMITA

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA KOMPUTER**

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRAK

EVALUASI KINERJA METODE SUPPORT VECTOR MACHINE (SVM), NAIVE BAYES DAN DECISION TREE UNTUK DIAGNOSA PENYAKIT JANTUNG

Oleh

SENDY HANI PRAMITA

Jantung adalah otot yang bekerja paling keras di tubuh. Jantung rata-rata berdetak 100.000 kali sehari, untuk memasok oksigen ke seluruh tubuh. Ada berbagai jenis penyakit jantung. Pembelajaran mesin paling banyak digunakan untuk prediksi penyakit di bidang medis. Banyak peneliti menjadi tertarik menggunakan pembelajaran mesin untuk mendiagnosis penyakit karena membantu mengurangi waktu diagnosis dan meningkatkan akurasi dan efisiensi. Data yang digunakan pada penelitian ini yaitu data dari *Heart Disease Dataset* yang bersumber dari *Public Health Dataset* dengan jumlah data 1025 dan atribut 14 atribut. Metode yang digunakan dalam klasifikasi penderita penyakit jantung yaitu Support Vector Machine (SVM), Naive Bayes dan *Decision Tree*. Mendapatkan hasil perbandingan hasil klasifikasi penyakit jantung model *Support Vector Machine* (SVM), model Naive Bayes dan model *Decision Tree* dapat digunakan sebagai salah satu metode prediksi penyakit jantung. *Model Support Vector Machine* (SVM) dan *Decision Tree* memiliki performa ketepatan prediksi sangat baik dengan memperoleh nilai akurasi sebesar 99% dan metode Naive Bayes yang memperoleh nilai akurasi yaitu 84%.

Kata Kunci : Jantung, Pembelajaran Mesin, Klasifikasi.

ABSTRACT

PERFORMANCE EVALUATION OF SUPPORT VECTOR MACHINE (SVM), NAIVE BAYES AND DECISION TREE METHODS FOR HEART DISEASE DIAGNOSIS

By

SENDY HANI PRAMITA

The heart is the hardest working muscle in the body. The average heart beats 100,000 times a day, to supply oxygen throughout the body. There are different types of heart disease. Machine learning is most widely used for disease prediction in the medical field. Many researchers are becoming interested in using machine learning to diagnose diseases because it helps reduce diagnosis time and increases accuracy and efficiency. The data used in this study are data from the Heart Disease Dataset which originates from the Public Health Dataset with a total of 1025 data and 14 attributes. The methods used in the classification of people with heart disease are Support Vector Machine (SVM), Naive Bayes, and Decision Tree. Obtaining the results of a comparison of the results of the classification of heart disease with the Support Vector Machine (SVM) model, the Naive Bayes model, and the Decision Tree model can be used as a method of predicting heart disease. The Support Vector Machine (SVM) and Decision Tree models have very good predictive accuracy performance by obtaining an accuracy value of 99% and the Naive Bayes method which obtains an accuracy value of 84%.

Keywords : Heart, Machine Learning, Classification.

Judul Skripsi : **EVALUASI KINERJA METODE SUPPORT VECTOR MACHINE (SVM), NAIVE BAYES DAN DECISION TREE UNTUK DIAGNOSA PENYAKIT JANTUNG**

Nama Mahasiswa : **Sendy Hani Pramita**

Nomor Pokok Mahasiswa : 1917051030


Program Studi : S1 Ilmu Komputer

Jurusan : Ilmu Komputer

Fakultas : Matematika dan Ilmu Pengetahuan Alam

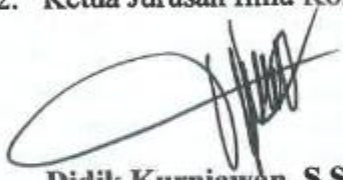
MENYETUJUI

1. Komisi Pembimbing


Prof. Admi Syarif, Ph.D.
NIP. 19670103 199203 1 003


Rahman Taufik, S.Pd, M. Kom
NIP. 19930627 202203 1 007

2. Ketua Jurusan Ilmu Komputer


Didik Kurniawan, S.Si., M.T.
NIP. 19800419 200501 1 004


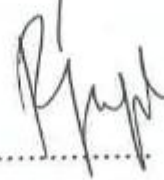

MENGESAHKAN

1. Tim Penguji

Ketua : **Prof. Admi Syarif, Ph.D**

Penguji I : **Rahman Taufik, S.Pd, M. Kom.**

Penguji II : **Rico Andrian, S.Si., M.Kom.**


.....

.....

.....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi : 23 Agustus 2023

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Sendy Hani Pramita

NPM : 1917051030

Dengan ini menyatakan bahwa skripsi saya yang berjudul “Evaluasi Kinerja Metode *Support Vector Machine* (SVM), Naive Bayes dan *Decision Tree* Untuk Diagnosa Penyakit Jantung” merupakan karya saya sendiri bukan karya orang lain. Semua tulisan yang tertulis dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Jika di kemudian hari terbukti bahwa skripsi saya merupakan hasil penjiplakan karya orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang telah saya peroleh.

Bandar Lampung, 23 Agustus 2023

Penulis



Sendy Hani Pramita

NPM. 1917051030

RIWAYAT HIDUP



Penulis lahir di Metro pada tanggal 11 Oktober 2000, sebagai anak ketiga dari empat bersaudara dari Alm. Bapak Suyanto dan Ibu Suharti. Penulis menyelesaikan pendidikan formal di Taman Kanak-kanak Aisyiyah Busthanul Athfal Adipuro tahun 2006, kemudian melanjutkan ke SDN 1 Adipuro tahun 2012, lalu SMP Negeri 3 Metro tahun 2015, dan terakhir SMA Negeri 2 Metro tahun 2018.

Penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN pada tahun 2019. Penulis telah mengikuti beberapa kegiatan sejak menjadi mahasiswa, seperti tercantum di bawah ini.

1. Menjadi anggota Printer Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2019/2020.
2. Menjadi anggota Bidang Kaderisasi Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2020/2021.
3. Melaksanakan Karya Wisata Ilmiah (KWI) di Desa Tambah Dadi, Kec Purbolinggo, Kabupaten Lampung Timur, pada bulan Desember tahun 2019.
4. Menjadi Staff Ahli BEM U KBM UNILA periode 2020/2021.
5. Menjadi Sekretaris Pelaksana Program Orientasi Jurusan Ilmu Komputer (PRINTER) 2020.
6. Menjadi Bendahara Bidang Internal Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2021/2022.

7. Melaksanakan Kerja Praktik di PT Bank Rakyat Indonesia (Persero) Tbk. Cabang Tanjung Karang pada periode 1 tahun 2022.
8. Melaksanakan KKN di Desa Mekar Jaya, Kecamatan Jabung, Kabupaten Lampung Timur pada periode II tahun 2022

MOTTO

The best and the beautiful in the world cannot be seen or even touched. They must be felt with the heart

(Helen Keller)

Have the courage to follow your heart and intuition. They somehow already know what you truly want to become.

(Steve Jobs)

Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum, sebelum mereka mengubah keadaan diri mereka sendiri.

(QS Ar Rad 11)

PERSEMBAHAN

Puji syukur yang tak terhingga saya ucapkan kepada Allah SWT atas segala Rahmat dan Karunia-Nya serta shalawat dan salam senantiasa tercurahkan Kepada Nabi Muhammad SAW sehingga saya menyelesaikan skripsi ini.

Kupersembahkan karya ini kepada:

Alm. Papa Terhebat dan Mama Tercinta, Terima kasih atas segala pengorbanan, perjuangan, kesabaran dan kasih sayang yang telah kalian berikan selama ini dalam membesarkan, mendidik, mendoakan, mendukung dan memberikan semangat dalam setiap langkah perjalanan yang telah saya lalui untuk mencapainya. keberkahan di dunia dan di akhirat. Kepada kedua kakakku dan support systemku, Renni dan Anggi yang aku sayangi, aku ucapkan terima kasih.

Keluarga Besar Ilmu Komputer 2019,

Serta Almamater Tercinta, Universitas Lampung.

SANWACANA

Segala puji atas kehadiran Allah SWT atas limpahan berkah, rahmat, dan karunia-Nya, shalawat serta salam yang senantiasa tercurahkan kepada junjungan kita Nabi Muhammad SAW. Sehingga penulis dapat menyelesaikan penulisan skripsi ini dengan tepat waktu. Skripsi ini disusun sebagai syarat untuk memperoleh gelar Sarjana Komputer di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung dengan judul **“Evaluasi Kinerja Metode *Support Vector Machine* (SVM), Naive Bayes dan *Decision Tree* Untuk Diagnosa Penyakit Jantung”**.

Terima kasih penulis ucapkan kepada semua pihak yang telah membantu dan berperan besar dalam menyusun skripsi ini, antara lain.

1. Allah SWT yang telah memberikan limpahan berkah, rahmat, hidayah, dan karunia-nya.
2. Kedua orang tuaku tercinta, Alm. Papa Suyanto dan Mama Suharti serta kedua kakakku Renni Wulandari dan Anggi Rejita, yang selalu memberikan doa, dukungan, semangat, motivasi, dan kasih sayang luar biasa tak terhingga. Semoga Allah SWT selalu memberikan kebahagiaan dan keberkahan dalam kehidupan kalian di dunia dan di akhirat.
3. Prof. Admi Syarif, Ph.D. selaku dosen pembimbing utama yang telah meluangkan banyak waktu, memberikan masukan dan saran hingga motivasi yang bermanfaat guna menyempurnakan penulisan skripsi ini.
4. Bapak Rahman Taufik, S.Pd, M.Kom. selaku dosen pembimbing kedua yang telah memberikan masukan dan saran yang bermanfaat guna menyempurnakan penulisan skripsi ini.

5. Bapak Rico Andrian, S.Si., M.Kom. selaku dosen pembahas yang telah memberikan masukan dan saran yang bermanfaat guna menyempurnakan penulisan skripsi ini.
6. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan FMIPA Universitas Lampung.
7. Bapak Didik Kurniawan, S.Si., M.T. selaku Ketua Jurusan Ilmu Komputer Universitas Lampung.
8. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. selaku Sekretaris Jurusan Ilmu Komputer Universitas Lampung.
9. Bapak dan Ibu Dosen Jurusan Ilmu Komputer yang telah memberikan banyak ilmu dan pengalaman selama penulis menjalani perkuliahan.
10. Ibu Nora, Bang Zai, dan Mas Noval yang sudah membantu urusan administrasi bagi penulis di Jurusan Ilmu Komputer.
11. Sherly Monica, A.Md.Kes. dan Nurul Afrisca, S.PWK. sahabat yang tidak pernah bosan mendengarkan keluh kesah, memberikan dukungan, menemani dan memberikan semangat untuk dapat menyelesaikan skripsi ini. Terima kasih untuk apapun yang sudah diberikan dari awal bertemu di 2012.
12. Dias Dinar Mozza Nabila, satu-satunya adik sepupu terbaik yang mendukung dan banyak membantu pada skripsi ini.
13. Grup Hangout, Zahara Liza Mulyani, Devi Ramdhia Fitri, Salsabilla Julia Farhana, Olivia Desti Riana, Hani Cita Lestari, Finka Marisa Genanda Sufie dan Vira Verina. sahabat- sahabatku yang sejak awal kuliah di jurusan ilmu komputer yang selalu ada dan saling membantu satu sama lain, tempatku menuangkan segala kebahagiaan dan kesedihan. Terima kasih untuk segalanya dari awal ada di Jurusan Ilmu Komputer.
14. Keluarga Besar Ilmu Komputer 2019 yang tidak bisa disebut satu persatu yang sudah bersedia menjadi teman kelompok, teman diskusi, hingga menjadi teman berbagi cerita. Terima kasih sudah memberikan banyak cerita dan pengalaman selama menempuh perkuliahan di Jurusan Ilmu Komputer Universitas Lampung.

DAFTAR ISI

	Halaman
DAFTAR ISI	i
DAFTAR TABEL	iii
DAFTAR GAMBAR	vi
I. PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian.....	6
II. TINJAUAN PUSTAKA	
2.1. Penelitian Terdahulu.....	7
2.2. Tinjauan Pustaka	11
2.2.1. Jantung	11
2.2.2. Penyakit Jantung	12
2.2.3. Jenis-jenis Penyakit Jantung dan Gejalanya	13
2.2.4. <i>Machine Learning</i>	14
2.2.5. <i>Exploratory Data Analysis (EDA)</i>	15
2.2.6. <i>Preprocessing</i>	15
2.2.7. Support Vector Machine (SVM).....	16
2.2.8. Naive Bayes	19
2.2.9. <i>Decision Tree</i>	20
2.2.10. <i>K-Fold Cross Validation</i>	22

2.2.11. <i>Hyperparameter Tuning</i>	22
2.2.12. <i>Confusion Matrix</i>	23

III. METODOLOGI PENELITIAN

3.1. Tempat dan Waktu Penelitian	27
3.2. Dataset dan Perangkat Penelitian	27
3.3. Alur Kerja Penelitian	30
3.4. Studi Literatur	31
3.5. Pengumpulan Data	31
3.6. <i>Preprocessing</i>	31
3.7. Penerapan Algoritma SVM, Naive Bayes dan <i>Decision Tree</i>	32
3.8. Hasil	35

IV. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data	36
4.2. Tahapan Percobaan	36
4.2.1. <i>Import Data</i>	36
4.2.2. <i>Exploratory Data Analysis (EDA)</i>	36
4.2.3. <i>Preprocessing</i>	48
4.2.4. Pembagian Data	48
4.2.5. Klasifikasi <i>Support Vector Machine (SVM)</i>	49
4.2.6. Klasifikasi Naive Bayes	55
4.2.7. Klasifikasi <i>Decision Tree</i>	59
4.3. Pembahasan	65
4.3.1. Hasil Perbandingan	67

V. KESIMPULAN DAN SARAN

5.1. Kesimpulan	71
5.2. Saran	71

DAFTAR PUSTAKA

DAFTAR TABEL

Tabel 1 Penelitian Terdahulu	7
Tabel 2. Jenis-jenis Penyakit Jantung dan gejalanya.	13
Tabel 3 Jenis Kernel <i>Support Vector Machine</i> (SVM)	18
Tabel 4. Deskripsi Paramater yang Digunakan.....	23
Tabel 5 <i>Confusion Matrix</i>	24
Tabel 6 Deskripsi Atribut Penelitian.....	27
Tabel 7. Iterasi 5 <i>k-fold cross-validation</i> sebelum <i>hyperparameter tuning</i> pada SVM	49
Tabel 8. <i>Confusion matrix</i> SVM sebelum <i>hyperparameter tuning</i>	50
Tabel 9. Hasil kinerja metode SVM sebelum <i>hyperparameter tuning</i>	50
Tabel 10. Parameter pada SVM sebelum <i>hyperparameter tuning</i>	51
Tabel 11. Iterasi 5 <i>k-fold cross-validation</i> metode SVM sesudah <i>hyperparameter tuning</i>	53
Tabel 12. <i>Confusion matrix</i> SVM sesudah <i>hyperparameter tuning</i>	53
Tabel 13. Hasil kinerja metode SVM sesudah <i>hyperparameter tuning</i>	54
Tabel 14. Iterasi 5 <i>k-fold cross-validation</i> metode Naive Bayes sebelum <i>hyperparameter tuning</i>	55
Tabel 15. <i>Confusion matrix</i> Naive Bayes sebelum <i>hyperparameter tuning</i>	55
Tabel 16. Hasil kinerja metode Naive Bayes sebelum <i>hyperparameter tuning</i> ...	56
Tabel 17. Parameter pada Naive Bayes.....	57
Tabel 18. Iterasi <i>k-fold cross-validation</i> metode Naive Bayes sesudah <i>hyperparameter tuning</i>	58
Tabel 19. <i>Confusion matrix</i> Naive Bayes sesudah <i>hyperparameter tuning</i>	58
Tabel 20. Hasil kinerja metode Naive Bayes sesudah <i>hyperparameter tuning</i>	59
Tabel 21. Iterasi <i>k-fold cross-validation</i> metode <i>Decision Tree</i> sebelum <i>hyperparameter tuning</i>	60

Tabel 22. <i>Confusion matrix Decision Tree</i> sebelum <i>hyperparamater tuning</i>	60
Tabel 23. Hasil kinerja metode <i>Decision Tree</i> sebelum <i>hyperparamater tuning</i> .	61
Tabel 24. Parameter pada <i>Decision Tree</i> sebelum <i>hyperparamater tuning</i>	62
Tabel 25. Iterasi <i>k-fold cross-validation</i> metode <i>Decision Tree</i> sesudah <i>hyperparamater tuning</i>	63
Tabel 26. <i>Confusion matrix Decision Tree</i> sesudah <i>hyperparamater tuning</i>	64
Tabel 27. Hasil kinerja metode <i>Decision Tree</i> sesudah <i>hyperparamater tuning</i> ..	64
Tabel 28. Hasil Kinerja Metode SVM	65
Tabel 29. Hasil Kinerja Metode Naive Bayes.....	66
Tabel 30. Hasil Kinerja Metode <i>Decision Tree</i>	66
Tabel 31. Perbandingan kinerja metode SVM, Naive Bayes dan <i>Decision Tree</i> .	69
Tabel 32. Perbandingan Hasil Akurasi	69

DAFTAR GAMBAR

Gambar 1 SVM berusaha mencari hyperplane terbaik	17
Gambar 2. Model susunan <i>Decision Tree</i>	20
Gambar 3. Alur Penelitian.....	30
Gambar 4. <i>Pie Chart</i> Target.....	38
Gambar 5. Rasio jenis kelamin	39
Gambar 6. Proporsi target untuk jenis kelamin.....	39
Gambar 7. Faktor risiko gula darah puasa dengan membandingkan <i>Sex</i> dan <i>Target</i>	40
Gambar 8. Distribusi dari tiga faktor risiko	41
Gambar 9. Distribusi faktor risiko untuk setiap jenis kelamin.....	41
Gambar 10. Grafik histogram jumlah nilai pada fitur.....	42
Gambar 11. Grafik batang proporsi jenis nyeri dada (i)	43
Gambar 12. Grafik batang proporsi jenis nyeri dada (ii)	43
Gambar 13. Grafik batang jumlah pasien dengan hasil EKG(i)	44
Gambar 14. Grafik batang jumlah pasien dengan hasil EKG (ii)	45
Gambar 15. Grafik distribusi dan <i>box plot</i> dari <i>ST-depression</i>	45
Gambar 16. Grafik sebaran data dari beberapa fitur numerik.....	46
Gambar 17. Plot hubungan antara beberapa fitur numerik	47
Gambar 18. <i>Heatmap</i> korelasi antar fitur.....	48

I. PENDAHULUAN

1.1.Latar Belakang

Jantung adalah otot yang bekerja paling keras di tubuh. Jantung rata-rata berdetak 100.000 kali sehari, untuk memasok oksigen ke seluruh tubuh. Darah yang dipompa oleh jantung juga mengangkut limbah produk seperti karbon dioksida untuk dikeluarkan dari tubuh. Fungsi jantung yang tepat sangat penting untuk mendukung kehidupan. Karena manusia tidak dapat meningkatkan atau menurunkan kecepatan denyut jantung saat jantung memompa darah atau bekerja, jantung tidak pernah berhenti bekerja saat manusia sedang istirahat. Karena itu, jika manusia dapat mengontrol keadaan dan denyut jantung mereka sendiri, mereka akan kewalahan dalam mengatur jantung mereka sendiri, yang pada gilirannya akan mengakibatkan kematian. (Adi, 2022)

Ada berbagai jenis penyakit jantung seperti penyakit katup jantung, Penyakit katup jantung adalah keadaan di mana satu atau lebih katup jantung tidak dapat berfungsi normal. Yang kedua adalah aritmia, yang merupakan gangguan pada irama jantung. Yang ketiga adalah endokarditis, yang merupakan infeksi pada lapisan dalam jantung. Yang keempat adalah penyakit jantung koroner, yang merupakan kerusakan pada jantung akibat gangguan pada arteri koroner, yang bertanggung jawab untuk menyediakan jantung dengan nutrisi, oksigen, dan darah. Yang kelima, gagal jantung, ditandai dengan jantung yang tidak efisien memompa darah ke seluruh tubuh. Yang keenam, kardiomiopati, adalah kondisi di mana bilik jantung melebar, yang menyebabkan otot jantung menjadi lemah dan tidak dapat memompa darah dengan baik. Dan yang ketujuh yaitu, serangan jantung biasanya terjadi ketika gumpalan darah mencegah aliran darah masuk ke jantung. (Dona et al., 2021)

Berdasarkan *American Heart Association* (2014) ada sejumlah penyebab penyakit jantung, termasuk diet yang tidak sehat yang mengandung lemak jenuh dan kolesterol. Selanjutnya, penyakit jantung dapat disebabkan oleh kurangnya aktivitas fisik. Selain itu, obesitas, alkohol, merokok, tekanan darah tinggi, kolestrol tinggi, diabetes melitus, genetika, dan riwayat keluarga dapat menjadi penyebab penyakit jantung karena faktor genetik dapat mewariskan kelainan tekanan darah tinggi, penyakit jantung, dan kondisi terkait lainnya. Selain itu, risiko penyakit jantung meningkat dengan usia. Hal ini wajar karena fungsi organ tubuh manusia menurun dengan usia.

Penyakit arteri koroner (*Coronary artery disease/CAD*), umumnya dikenal sebagai penyakit jantung, adalah kondisi dimana kolesterol, kalsium, dan lainnya menumpuk di arteri yang memasok darah ke jantung. Bahan ini mengeras membentuk plak yang menghalangi aliran darah ke jantung. Ketika koroner arteri menyempit karena penumpukan plak atau lainnya. Menyebabkan, otot jantung kekurangan oksigen dan mengalami nyeri dada yang dikenal sebagai angina.

Dalam kehidupan kita sehari-hari, generasi saat ini sangat sibuk dengan kita jadwal rutin harian yang terasa gugup, kegelisahan dan stres. Setiap individu memiliki denyut nadi yang khas tingkat dan tekanan darah yang berkisar 60-100 BPM untuk denyut nadi dan 120/80 sampai 140/90 untuk tekanan darah. Di dunia luas, penyakit jantung adalah masalah utama pada manusia kehidupan. Jantung berarti "Kardio". Kategori penyakit jantung disebut Penyakit kardiologis.

Selain penyakit arteri koroner (*Coronary artery disease/CAD*), penyakit lain yang terjadi di jantung adalah penyakit gagal jantung. Gagal jantung adalah suatu kondisi dimana jantung tidak mampu memasok darah yang cukup untuk memenuhi kebutuhan tubuh. Arteri koroner sebagai bagian integral dari jantung bertanggung jawab untuk memasok darah ke jantung. Penyakit arteri koroner (penyempitan atau penyumbatan arteri) adalah jenis penyakit jantung yang paling umum dan penyebab gagal jantung yang paling umum.

Ada banyak kondisi berbahaya yang mengakibatkan penyakit gagal jantung. Kondisi ini dapat dimasukkan ke dalam dua kategori, dengan kategori pertama terdiri dari kondisi berisiko atau berbahaya yang tidak dapat diubah, misalnya jenis kelamin pasien, usia, dan riwayat keluarga. Kategori kedua, yang dapat diubah, terdiri dari kondisi yang dikaitkan dengan cara hidup pasien, misalnya kebiasaan merokok, kadar kolesterol tinggi, tekanan darah tinggi, dan kurangnya aktivitas fisik. Selain itu, gejala gagal jantung yang lazim termasuk dispnea (sesak napas), edema (kaki bengkak), kelelahan, dan kelemahan.

Kebanyakan pasien penderita penyakit jantung tidak mengetahui gejala-gejala awal yang dirasakan dan tidak sedikit banyak penderita penyakit jantung koroner yang meninggal disebabkan oleh serangan jantung. Masih kurangnya kesadaran terhadap pola hidup yang sehat dan kurangnya informasi penyakit jantung koroner yang dapat membuat seseorang tidak dapat mengenali gejala awalnya. Proses untuk mendeteksi penyakit jantung dapat dilakukan dengan cara manual, yakni dengan konsultasi langsung ke dokter spesialis jantung dan melakukan beberapa pemeriksaan laboratorium yang kemudian harus dikonsultasikan kembali oleh dokter spesialis jantung. Hal ini tentu saja memerlukan biaya yang relatif besar. Dengan resiko kematian yang sangat tinggi, maka diperlukan suatu sistem yang dapat mendeteksi penyakit jantung koroner pada penderita secara akurat serta dengan biaya yang tidak besar. (Wibisono & Fahrurrozi, 2019)

Penyakit kardiovaskular adalah penyebab utama kematian secara global. Setiap tahun 17.9 juta nyawa diambil oleh penyakit ini (WHO, 2021). Dan berdasarkan data Riset Kesehatan Dasar (Riskesdas) tahun 2018, angka kejadian penyakit jantung dan pembuluh darah semakin meningkat dari tahun ke tahun. Setidaknya, 15 dari 1000 orang, atau sekitar 2.784.064 individu di Indonesia menderita penyakit jantung. (Nurmasani & Pristyanto, 2021) Penyakit jantung dapat dicegah. Jika diagnosa dilakukan dengan tepat, kita dapat mengurangi risiko masalah jantung. Penyakit jantung bisa dicegah dengan deteksi dini. Tentunya dengan kemajuan teknologi angka tersebut dapat diminimalisir. Salah satu teknologi yang ada sekarang yaitu *machine learning*

yang dapat digunakan untuk mendeteksi penyakit jantung. Algoritma *machine learning* memainkan peran penting dalam menganalisis data pelayanan kesehatan.

Hal ini mendorong banyak penelitian terhadap penyakit jantung koroner, salah satunya menggunakan metode berbasis komputer. Metode ini banyak dikembangkan dengan bantuan komputasi cerdas yang mampu mengolah data dalam jumlah yang besar. Pengolahan data dalam jumlah besar dapat dilakukan dengan klasifikasi menggunakan algoritma tertentu sehingga hasilnya cepat dan akurat.

Pembelajaran mesin paling banyak digunakan untuk prediksi penyakit di bidang medis. Banyak peneliti menjadi tertarik menggunakan pembelajaran mesin untuk mendiagnosis penyakit karena membantu mengurangi waktu diagnosis dan meningkatkan akurasi dan efisiensi. Beberapa penyakit dapat didiagnosis dengan menggunakan teknik pembelajaran mesin, namun fokus penelitian ini akan tertuju pada diagnosa penyakit jantung. Karena penyakit jantung adalah penyebab utama dari kematian di dunia saat ini.

Naive Bayes, *Decision Tree* dan SVM merupakan metode algoritma data mining yang digunakan untuk melakukan klasifikasi. Beberapa penelitian yang sebelumnya telah dilakukan, SVM, Naive Bayes dan *Decision Tree* diketahui memiliki akurasi yang baik. (Adi, 2022; Al-Janabi et al., 2018; Golpour et al., 2020; Maulidah et al., 2021; Supriyadi et al., 2022). Penelitian ini dilakukan dengan tujuan untuk melakukan perbandingan antara algoritma Naive Bayes, *Decision Tree* dan SVM dalam mendiagnosa penyakit jantung.

Berdasarkan latarbelakang tersebut maka dilakukan penelitian dengan melakukan evaluasi kinerja dari metode *Support Vector Machine* (SVM), Naive Bayes dan *Decision Tree* untuk mengetahui algoritma yang memiliki akurasi yang lebih tinggi dalam hal ini yaitu prediksi penyakit jantung karena kedua algoritma tersebut memiliki keunggulan dalam hal mengklasifikasi sebuah data. Sehingga diambil judul “Evaluasi Kinerja Metode *Support Vector*

Machine (SVM), Naive Bayes Dan *Decision Tree* Untuk Diagnosa Penyakit Jantung”.

1.2.Rumusan Masalah

Berdasarkan penjelasan pada latar belakang dapat ditarik rumusan masalah sebagai berikut:

1. Bagaimana model klasifikasi penyakit jantung menggunakan metode *Support Vector Machine* (SVM) ?
2. Bagaimana model klasifikasi penyakit jantung menggunakan metode Naïve Bayes ?
3. Bagaimana model klasifikasi penyakit jantung menggunakan metode *Decision Tree* ?
4. Bagaimana perbandingan hasil klasifikasi penyakit jantung model *Support Vector Machine* (SVM), model Naive Bayes dan model *Decision Tree* ?

1.3. Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan pada penelitian ini yaitu data dari *Heart Disease Dataset* yang bersumber dari *Public Health Dataset* yang dapat diakses melalui <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>.
2. Metode yang digunakan dalam klasifikasi penderita penyakit jantung yaitu *Support Vector Machine* (SVM), Naive Bayes dan *Decision Tree*.

1.4.Tujuan Penelitian

Tujuan dari penelitian ini antara lain adalah sebagai berikut :

1. Mendapatkan model klasifikasi penyakit jantung menggunakan metode *Support Vector Machine* (SVM)

2. Mendapatkan model klasifikasi penyakit jantung menggunakan Naïve Bayes
3. Mendapatkan model klasifikasi penyakit jantung menggunakan *Decision Tree*
4. Mendapatkan hasil perbandingan klasifikasi metode *Support Vector Machine* (SVM), metode Naïve Bayes dan metode *Decision Tree* serta membandingkan rata-rata akurasi klasifikasi dari ketiga metode tersebut.

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah memberikan informasi terkait hasil klasifikasi data menggunakan metode *Support Vector Machine* (SVM), *Naïve Bayes* dan *Decision Tree* dalam diagnosis penyakit jantung.

II. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Tabel 1 Penelitian Terdahulu

No	Penelitian	Data	Metode	Hasil
1	<i>Machine learning classification techniques for heart disease prediction</i> (Al-Janabi et al., 2018)	Dataset yang digunakan adalah dataset penyakit jantung diperoleh dari UCI (<i>University of California, Irvine C.A.</i>). Dataset ini berisi empat database dari empat rumah sakit. Dataset memiliki 14 atribut dengan jumlah data pasien 303 pasien.	Naive Bayes, ANN, <i>Decision Tree</i> , dan SVM	Metode Naive Bayes, akurasi : 86,42%, Metode ANN, akurasi : 99.25%, Metode <i>Decision Tree</i> akurasi ; 56.76%, Metode SVM, akurasi : 61.86%
2	<i>A Feature-Driven Decision Support System for Heart Failure Prediction</i>	Pada penelitian ini diambil 297 sampel data. <i>Dataset dari University of California.</i> Dataset memiliki 13 atribut.	Gaussian Naive Bayes	Metode Gaussian Naive Bayes, akurasi : 84,05% Sensitivitas 87,80 %

No	Penelitian	Data	Metode	Hasil
	<p><i>Based on χ^2 Statistical Model and Gaussian Naive Bayes</i> (Ali et al., 2019)</p>			<p>Spesifisitas : 97,95%.</p>
3	<p><i>Prediction of Heart Disease UCI Dataset Using Machine Learning Algorithms</i> (Anderies et al., 2022)</p>	<p>Dalam penelitian ini, data diperoleh dari UCI machine learning repository dengan menggunakan 14 atribut.</p>	<p>SVM, Naive Bayes, Regresi logistik, Neural Network, Decision Tree dan KNN</p>	<p>Accuracy, Precision, Recall, dan F1 Score</p> <p>SVM : 85%, 0.97, 0.79, 0.87</p> <p>Naive Bayes : 83.33%, 0.96, 0.76, dan 0.85</p> <p>Regresi Logistik : 83.33%, 0.94, 0.79, dan 0.86</p> <p>Neural Network: 80%, 0.91, 0.76 dan 0.83</p> <p>Decision Tree : 70%, 0.86, 0.63 dan 0.72</p>

No	Penelitian	Data	Metode	Hasil
				KNN 78%, 0.90 0.74, dan 0.81
4	<i>Analysis of Heart Disease Using Parallel and Sequential Ensemble Methods With Feature Selection Techniques</i> (Yadav & Pal, 2021)	Dalam penelitian ini, digunakan informasi terkait penyakit jantung dari repositori UCI. Dataset berisi 1025 data pasien dengan 14 atribut, pasien sakit dan tidak sakit dalam variabel target.	<i>Random Forest Ensemble</i>	Metode <i>random forest ensemble</i> , akurasi : 99%.
5	Komparasi Metode Support Vector Machine (SVM), <i>K-Nearest Neighbors</i> (KNN), dan <i>Random Forest</i> (RF) Untuk Prediksi Penyakit Gagal Jantung (Adi, 2022)	Pada penelitian ini dataset yang digunakan diambil dari laman UCI Machine Learning dengan judul <i>Heart Failure Clinical Records Dataset</i> . Dataset terdiri dari 299 data pasien penyakit gagal jantung dan masing-masing 13 atribut.	<i>Support Vector Machine</i> (SVM), <i>K-Nearest Neighbors</i> (KNN), dan <i>Random Forest</i> (RF)	Metode SVM, <i>Precesions</i> : 90%, <i>Recall</i> : 98%, <i>F1 Score</i> : 93%, <i>Accuracy</i> : 97%. Metode <i>Random Forest</i> , <i>Precesions</i> : 90%, <i>Recall</i> : 98%, <i>F1 Score</i> : 93%, <i>Accuracy</i> : 97%. Metode KNN <i>Precesions</i> : 96%,

No	Penelitian	Data	Metode	Hasil
				<i>Recall : 75%, F1 Score : 81%, Accuracy : 93%.</i>
6	<i>A machine learning approach: Using predictive analytics to identify and analyze high risks patients with heart disease (Khennou et al., 2019)</i>	Penelitian ini, dataset didapat dari <i>UCI Machine Learning Repository</i> yang bersumber dari Cleveland, Hungaria dan Swiss.	<i>Support Vector Machine</i>	Metode SVM, akurasi : 87%
7	A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms (Haq et al., 2018)	Dataset yang digunakan didapat dari dataset <i>University of California Irvine (UCI)</i> .	<i>Logistic Regression</i> dan <i>Support Vector Machine</i>	Metode <i>Logistic Regression</i> , accuracy : 89%, Metode <i>Support Vector Machine</i> , accuracy : 88%.
8	Prediksi Penyakit Jantung dengan Algoritma Klasifikasi (Science, 2019)	Dataset yang akan digunakan yaitu <i>Statlog Heart Disease Dataset</i> yang berasal dari <i>UCI machine learning repository</i>	Naive Bayes, SVM, dan, <i>Logistic Regression</i> .	Metode Naive Bayes, akurasi : 84.07%, Metode SVM, akurasi : 81.85%,

No	Penelitian	Data	Metode	Hasil
				Metode <i>Logistic Regression</i> , akurasi : 82.59%
9	Analisis Penyakit Jantung Koroner Menggunakan <i>Decision Tree</i> . (Akbar et al., 2019)	Kumpulan data diperoleh dari UCI <i>Heart Disease</i> untuk memprediksi apakah pasien menderita penyakit atau tidak.	<i>Decision Tree</i>	Metode <i>Decision Tree</i> , <i>accuracy</i> : 87.5% <i>Precision</i> : 87.2% <i>Recall</i> : 85% <i>F1-score</i> : 86.1%
10	<i>Diagnosis And Prediction Of Heart Disease Using Machine Learning Techniques</i> (Jeyaganesan et al., 2020)	Dataset bersumber dari <i>Public Health Dataset</i> pada <i>Kaggle</i> . Dataset terdiri dari 14 atribut. Total data pada dataset ini totalnya adalah 1025 data.	<i>Decision Tree</i> , <i>Logistic Regression</i> , <i>Random Forest</i> dan Naive Bayes	Metode <i>Decision Tree</i> : 84,5% <i>Logistic Regression</i> : 85,7% <i>Random Forest</i> : 93,7% Naive Bayes : 83,7%

2.2.Tinjauan Pustaka

2.2.1. Jantung

Jantung merupakan salah satu alat vital yang peranannya paling penting dalam tubuh manusia. Organ jantung mempunyai rongga yang berisi cairan

dimana fungsinya untuk melumasi jantung selama berdenyut dan jantung terletak di tengah tulang rusuk. Darah di dalam jantung dialirkan melalui bagian kanan dan kiri jantung itu sendiri. Jantung memiliki dua katup dimana katup tersebut berfungsi untuk memastikan bahwa darah hanya akan mengalir dalam satu arah saja.

Peran paling utama dari jantung yaitu untuk mensuplai O_2 atau oksigen ke bagian-bagian tubuh manusia dan mengeluarkan metabolit yang sudah tidak terpakai dari dalam tubuh. Jantung bekerja dengan cara menjadikan kumpulan darah yang sudah mengandung CO_2 atau karbondioksida dari seluruh tubuh lalu mengirimkannya ke paru-paru, kemudian di dalam paru-paru akan dilakukan proses perubahan dari karbondioksida menjadi oksigen. Lalu jantung akan mengirimkan kumpulan darah yang banyak mengandung oksigen dari paru-paru ke bagian-bagian tubuh manusia (Adi, 2022)

2.2.2. Penyakit Jantung

Penyakit jantung adalah suatu keadaan dimana jantung tidak dapat melaksanakan fungsinya dengan baik, sehingga kerja jantung sebagai pemompa darah dan oksigen ke seluruh tubuh terganggu. Terganggunya peredaran oksigen dan darah tersebut dapat disebabkan karena otot jantung yang melemah, adanya celah antara serambi kiri dan serambi kanan yang mengakibatkan darah bersih dan darah kotor bercampur (Naryadi, 2019)

Penyakit jantung biasanya terjadi karena kerusakan sel otot-otot jantung dalam memompa aliran darah keseluruh tubuh, yang disebabkan kekurangan oksigen yang dibawa darah ke pembuluh darah di jantung atau juga karena terjadi kejang pada otot jantung yang menyebabkan kegagalan organ jantung dalam memompa darah, sehingga menyebabkan kondisi jantung tidak dapat melaksanakan fungsinya dengan baik. Penyakit jantung dapat terjadi pada siapa saja di segala usia, jenis kelamin, pekerjaan, dan gaya hidup, selain itu penyakit jantung tidak bisa disembuhkan (Naryadi, 2019).

2.2.3. Jenis-jenis Penyakit Jantung dan Gejalanya

Tabel 2. Jenis-jenis Penyakit Jantung dan gejalanya.

Menurut WHO 2016 (Dona et al., 2021)

No	Jenis-jenis Penyakit Jantung	Gejala											
		Nyeri Dada	Sesak Napas	Keringat Dingin	Dada Berdebar	Mual	Hilang Kesadaran	Pusing	Cepat Lelah	Demam	Detak Jantung Tak Beraturan	Batuk	Bengkak pada Bagian Tubuh
1	Penyakit Jantung Koroner	✓	✓	✓	✓	✓							
2	Serangan Jantung	✓	✓			✓	✓						
3	Aritmia	✓	✓				✓	✓	✓	✓	✓		
4	Kardiomiopati	✓	✓						✓			✓	✓
5	Gagal Jantung	✓	✓						✓				
6	Penyakit Katup Jantung	✓	✓						✓		✓		
7	Endokarditis	✓	✓					✓		✓		✓	

2.2.4. *Machine Learning*

Machine learning adalah ilmu yang melibatkan pengembangan dan desain algoritma yang bertujuan untuk mengembangkan perilaku berdasarkan data empiris dalam program komputer. Pembelajaran mesin adalah ilmu yang bertujuan untuk memungkinkan mesin belajar sendiri atau berfungsi dari membangun model yang dibuat dari kumpulan data. Oleh karena itu, data sangat diperlukan ketika membangun sebuah model. Model yang dibuat biasanya digunakan pada mesin, memungkinkan mesin untuk belajar dan beroperasi sendiri. *Machine learning* adalah bidang kecerdasan buatan dalam sains yang bertujuan agar mesin atau komputer dapat melakukan pekerjaannya sendiri dengan menggunakan data yang ada. Data dapat digunakan untuk membuat aturan atau algoritma, memungkinkan mesin membuat keputusan sendiri berdasarkan aturan atau algoritma yang dibuat. *Machine learning* mempunyai 2 tipe teknik yaitu *supervised learning* dan *unsupervised learning*.

a. *Supervised Learning*

Supervised learning adalah salah satu tipe algoritma *machine learning* yang mempunyai dataset yang dikenal *training set* untuk membuat prediksi atau klasifikasi. Tujuan pembelajaran adalah membangun model yang dapat menghasilkan *output* yang benar untuk suatu data *input*. Contoh algoritma dalam *supervised learning* yaitu *Decision Tree*, KNN (*K-Nearest Neighbors*), *Logistic Regression*, SVM, dan Naive Bayes. Terdapat dua aplikasi utama dalam *machine learning* pada *supervised learning* yaitu klasifikasi dan regresi.

b. *Unsupervised Learning*

Unsupervised Learning adalah salah satu tipe algoritma *machine learning* yang digunakan untuk menarik kesimpulan dari *dataset* yang terdiri dari *input data labeled response*. Metode *unsupervised learning* yang paling umum adalah analisa cluster, yang digunakan pada analisa data untuk mencari pola-pola tersembunyi atau pengelompokan dalam data. Tujuan pembelajaran adalah membangun model yang dapat menemukan variabel

tersembunyi pada data pelatihan. Selanjutnya, variabel tersembunyi tersebut dapat digunakan untuk kebutuhan pemodelan (*latent variabel models*). Contoh *unsupervised learning* yaitu *K-Means clustering* dan *Hierarchical clustering*.

2.2.5. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah proses menganalisis dan menampilkan data bertujuan mendapatkan pemahaman yang lebih baik tentang wawasan dari data. Ada berbagai langkah yang dilakukan saat melakukan EDA, berikut ini adalah langkah - langkah umum yang dapat diambil dalam melakukan analisis EDA data:

- a. Memaksimalkan wawasan ke dalam kumpulan data.
- b. Mengungkap struktur data.
- c. Ekstrak variabel yang penting.
- d. Mendeteksi outlier dan anomali.
- e. Melakukan uji asumsi.
- f. Mengembangkan model.
- g. Menentukan faktor yang optimal.

2.2.6. Preprocessing

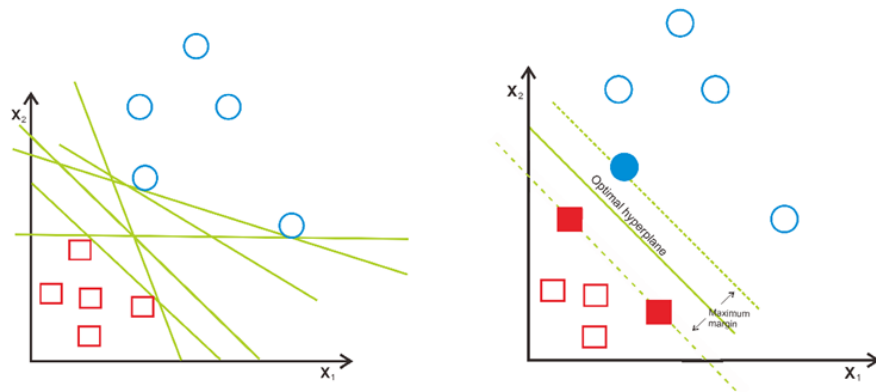
Preprocessing merupakan teknik awal data mining untuk mengubah data mentah atau biasa dikenal dengan raw data yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih dan bisa digunakan untuk pengolahan selanjutnya. Proses ini bisa juga disebut dengan langkah awal untuk mengambil semua informasi yang tersedia dengan cara membersihkan, memfilter, dan menggabungkan data-data tersebut. 3 masalah umum yang diselesaikan dalam tahap *preprocessing* adalah menangani *missing value*, *data noise*, dan data yang tidak konsisten. *Missing value* merupakan data yang tidak akurat karena informasi yang hilang menyebabkan informasi yang ada di dalamnya tidak relevan. *Missing*

value sering terjadi ketika ada masalah dalam proses pengumpulan, seperti kesalahan dalam *entry data* atau masalah dalam penggunaan biometrik. *Data noise* berisi data yang salah dan pencilan yang dapat ditemukan di kumpulan data. Pencilan dan data salah ini berisi informasi yang tidak berarti. Beberapa penyebab adanya *data noise* adalah karena kesalahan manusia berupa kesalahan pemberian label dan masalah lain selama pengumpulan data. Inkonsisten data terjadi ketika seseorang menyimpan file yang berisi data yang sama dengan format yang berbeda-beda. Beberapa inkonsisten data adalah duplikasi dalam format yang berbeda, kesalahan pada kode nama, dan lain sebagainya.

Preprocessing data sangat penting karena kesalahan, redundan, missing value, dan data yang tidak konsisten menyebabkan berkurangnya akurasi hasil analisis. Jadi, sebelum mengolah data, kita harus memastikan bahwa data yang akan kita gunakan merupakan data "bersih".

2.2.7. Support Vector Machine (SVM)

Support Vector Machine (SVM) mula-mula diperkenalkan oleh Vapnik pada tahun 1992 sebagai kumpulan proposional rancangan-rancangan yang unggul dalam bidang *pattern recognition*. Teori dasar *Support Vector Machine* (SVM) sudah ada puluhan tahun sebelumnya, seperti margin *hyperplane* (Duda & Hart tahun 1973, Cover tahun 1965, Vapnik 1964, dsb.), kernel diperkenalkan oleh Aronszajn tahun 1950). *Support Vector Machine* (SVM) adalah algoritme *supervised Machine learning* kerap dimanfaatkan untuk memprediksi dan pengklasifikasian. (Pisner & Schnyer, 2019).



Gambar 1 SVM berusaha mencari hyperplane terbaik

(Pisner & Schnyer, 2019)

Support Vector Machine (SVM) mempunyai teori dasar bahwa menemukan fungsi pemisah (*hyperplane*) secara optimal di antara fungsi yang ada untuk memisahkan dua macam kelas. Persamaan *hyperplane* diartikan baik ketika mempunyai margin terbesar. Margin adalah dua kali jarak antara *hyperplane* dan *support vector*, tepatnya *support vector* adalah titik yang berada paling dekat dengan *hyperplane* (Sari & Irhamah, 2020).

Gambar menjelaskan sebagian *pattern* yang merupakan anggota dari dua buah *class*: +1 dan -1. *Pattern* yang tergabung pada *class* -1 disimbolkan dengan warna merah (kotak), sedangkan *pattern* pada *class* +1, disimbolkan dengan warna kuning (lingkaran). Masalah klasifikasi dapat diterjemahkan dengan usaha menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) yang dapat dilihat pada *Hyperplane* pemisah terbaik antara kedua *class* dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* tersebut dengan *pattern* terdekat dari masing-masing *class*. *Pattern* yang paling dekat ini disebut sebagai *support vector* (Pisner & Schnyer, 2019).

Garis solid pada Gambar menunjukkan *hyperplane* yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua *class*, sedangkan titik merah dan kuning yang berada dalam lingkaran hitam adalah *support vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pembelajaran pada *Support Vector Machine* (SVM) (Pisner & Schnyer, 2019).

Pemodelan menggunakan metode *Support Vector Machine* (SVM) untuk menghasilkan akurasi yang terbaik. Dalam proses pemodelan ini menggunakan fungsi klasifikasi *Support Vector Machine* (SVM) untuk menjalankan algoritme pada python yaitu menggunakan *Support Vector Machine* (SVM) dengan mengimpor *Support Vector Classification* (SVC). *Support Vector Machine* (SVM) merupakan metode *supervised learning* yang digunakan untuk klasifikasi, regresi dan deteksi *outlier*.

Metode *Support Vector Machine* (SVM) dipilih karena terbukti memiliki akurasi yang baik untuk menganalisis teks. *Kernel* merupakan seperangkat fungsi matematika yang dimiliki *Support Vector Machine* (SVM). *Kernel* ini berfungsi untuk mengambil data sebagai input kemudian mengubah data tersebut menjadi ruang dimensi tingkat tinggi (*Kernel Space*). *Kernel* yang populer dalam *Support Vector Machine* (SVM) diantaranya *linear kernel*, *polynomial function*, *sigmoid*, *Gaussian radial basis function* (RBF) (Pisner & Schnyer, 2019). Syarat sebuah fungsi menjadi fungsi *kernel* adalah memenuhi Teorema Mercer yang menyatakan bahwa matriks *kernel* yang dihasilkan harus bersifat positif semi-definite (Muis et al., 2015). Jenis-jenis *kernel Support Vector Machine* (SVM) dapat dilihat pada tabel berikut:

Tabel 3 Jenis Kernel *Support Vector Machine* (SVM)

Jenis <i>kernel</i>	Definisi
Linear	$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i^t \vec{x}_j$

Jenis <i>kernel</i>	Definisi
Polynomial	$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i, \vec{x}_j + 1)^p$
Gaussian	$K(\vec{x}_i, \vec{x}_j) = \exp - \frac{ \vec{x}_i - \vec{x}_j }{2\sigma^2}$
Sigmoid	$K(\vec{x}_i, \vec{x}_j) = \tanh(a\vec{x}_i, \vec{x}_j + \beta)$

Data yang tidak dapat diklasifikasikan secara linear dapat menggunakan metode dengan cara mentransformasikan data kedalam ruang fitur (*feature space*). *Feature space* biasanya memiliki dimensi yang lebih tinggi dari vektor input (*input space*). Hal ini mengakibatkan komputasi pada *feature space* sangat besar jumlah *feature* yang tidak terhingga dan juga sulit mengetahui fungsi transformasi yang tepat. Untuk mengatasi masalah tersebut, pada *Support Vector Machine* (SVM) menggunakan *kernel trick* (Muis et al., 2015). Pada penelitian ini menggunakan metode *Support Vector Machine* dengan (SVM) *kernel* yang digunakan adalah *kernel* linear untuk pengolahan data, hasil komposisi *dataset* yang berbeda digunakan untuk mencari nilai akurasi yang terbaik.

2.2.8. Naive Bayes

Naive bayes adalah metode klasifikasi yang berdasarkan probabilitas dan Teorema Bayesian dengan asumsi bahwa setiap variabel X bersifat bebas atau berdiri sendiri dan tidak ada kaitannya dengan variabel lainnya. Naive Bayes termasuk ke dalam pembelajaran supervised, sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan untuk dapat mengambil keputusan. Pada tahapan pengklasifikasian akan dihitung nilai probabilitas dari masing-masing label kelas yang ada terhadap masukan yang diberikan. Algoritma Naive Bayes memberikan suatu cara mengkombinasikan peluang terdahulu dengan syarat kemungkinan menjadi sebuah formula yang dapat digunakan untuk menghitung peluang dari setiap kemungkinan yang terjadi (Arifin & Sasongko, 2018). Penghitungan pada Naive Bayes dapat dirumuskan seperti pada persamaan :

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Keterangan :

X : Data dengan class yang belum diketahui.

H : Hipotesis data X merupakan suatu class spesifik

$(H|X)$: Probabilitas hipotesis H berdasarkan kondisi

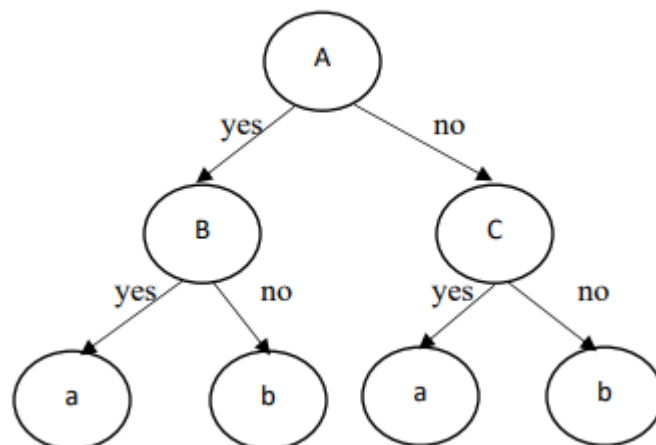
$X(H)$: Probabilitas hipotesis H (prior probability)

$(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

(X) : Probabilitas X

2.2.9. *Decision Tree*

Metode *Decision tree* merupakan salah satu metode teknik klasifikasi yang ada pada data mining. *Decision tree* adalah algoritma yang paling banyak digunakan untuk masalah klasifikasi. Metode *Decision Tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan. Pohon keputusan juga berguna untuk mengeksplorasi data dan menemukan hubungan tersembunyi antara sekumpulan kandidat variabel input dan variabel target. *Decision tree* terdiri dari beberapa simpul *tree's root*, *internal node* dan *leafs*.



Gambar 2. Model susunan *Decision Tree*

Konsep entropi digunakan untuk menentukan atribut mana yang dipisahkan oleh pohon. Dalam pohon keputusan, setiap simpul internal membagi ruang menjadi dua bagian atau lebih sesuai dengan fungsi diskrit dari nilai atribut input. Dalam kasus yang paling sederhana dan umum, setiap pengujian menganggap semua atribut tunggal, sehingga ruang di partisi kosong disesuaikan dengan nilai atribut. Klasifikasi menggunakan pohon keputusan dilakukan dengan melakukan *routing* dari simpul akar hingga mencapai simpul daun. Algoritma pohon keputusan meliputi ID3, C4.5, dan CART (Wibawa et al., 2018).

Algoritma C4.5 diperkenalkan oleh Quinlan sebagai versi perbaikan dari ID3. Dalam ID3, induksi decision tree hanya dapat dilakukan pada jalur bertipe kategorikal (nominal/ordinal), sedangkan tipe numerik (internal/rasio) tidak dapat menangani fitur yang melibatkan tipe numerik, melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (*deriving*) *rule set*. Algoritma C4.5 juga menggunakan kriteria gain dalam menemukan fitur yang menjadi pemecah simpul untuk pohon yang diinduksi (Muslim et al., 2019).

Ada beberapa langkah dalam algoritma untuk membangun pohon keputusan C4.5, Larose yaitu (Bahri & Lubis, 2020):

1. Menyiapkan data training. Data training diambil dari data histori yang pernah terjadi sebelumnya dan telah dikelompokkan pada kelas-kelas tertentu.
2. Hitung akar dari pohon. Akar diambil dari atribut yang terpilih, dengan menghitung nilai gain dari setiap atribut, nilai gain tertinggi menjadi akar pertama. Untuk menghitung nilai gain dari atribut, hitung terlebih dahulu nilai entropy.

Rumus Nilai Entropy:

$$entropy (s) = \sum_e = -p_i \times \log_2 p_i$$

Keterangan :

S : Himpunan Kasus

A : Fitur

N : Jumlah Partisi S

Pi : Proporsi dari Si terhadap S

3. Menghitung nilai *Gain* dengan rumus sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

4. Ulangi langkah 2 dan 3 hingga semua terpartisi.
5. Proses partisi pohon keputusan berakhir ketika:
- Semua record dari node N diberi kelas yang sama
 - Tidak ada atribut lain dalam record yang dipartisi lagi. Dalam cabang kosong tidak ada record di dalamnya.

2.2.10. *K-Fold Cross Validation*

Cross-validation atau yang bisa disebut dengan estimasi rotasi adalah teknik validasi model untuk menilai bagaimana hasil analisis statistik akan digeneralisasikan ke kumpulan data independen. Teknik ini terutama digunakan untuk membuat prediksi model dan memperkirakan seberapa akurat model prediksi saat dijalankan dalam praktik. Salah satu teknik validasi silang adalah validasi silang k-fold, yang memecah data menjadi k subset dari kumpulan data dengan ukuran yang sama. Penggunaan validasi silang k-fold untuk menghilangkan bias dalam data. Pelatihan dan pengujian dilakukan k kali. Pada percobaan pertama, subset S1 diperlakukan sebagai data uji dan subset lainnya diperlakukan sebagai data pelatihan, pada percobaan kedua subset S1, S3,...Sk menjadi data pelatihan dan S2 menjadi data uji, dan seterusnya.

2.2.11. *Hyperparameter Tuning*

Hyperparameter tuning memiliki peran yang sangat penting dalam mengoptimalkan kinerja algoritma *machine learning* (ML). Nilai *hyperparameter* tidak dapat ditentukan dari data dan mengambil seperti

yang diberikan saat mendefinisikan model, dengan kata lain, nilai *hyperparameter* harus ditentukan sebelum model menjalani proses pembelajaran. *Hyperparameter* adalah variabel yang mempengaruhi *output* model.

Tabel 4. Deskripsi Paramater yang Digunakan

Nama	Deskripsi
C	Nilai parameter C dalam SVM mengontrol trade-off antara kompleksitas model dan tingkat toleransi kesalahan.
gamma	Parameter yang digunakan oleh fungsi kernel RBF untuk pemetaan data ke ruang fitur yang dapat ditentukan nilai parameternya
<i>var_smoothing</i>	Untuk menangani masalah probabilitas nol atau sangat kecil.
<i>criterion</i>	Parameter ini mengacu pada kriteria yang digunakan untuk melakukan pemisahan atau pembagian pada setiap simpul dalam pohon keputusan. Beberapa criteria diantaranya : <ul style="list-style-type: none"> • <i>Gain</i> : criteria ini melakukan penyesuaian informasi untuk setiap atribut untuk memungkinkan luas dan keseragaman atribut. • <i>Gini</i> : criteria ini mengukur ketidaksetaraan antara distribusi karakteristik label. Pemisahan pada Atribut yang dipilih menghasilkan pengurangan indeks gini rata-rata dari himpunan bagian yang dihasilkan
<i>max_depth</i>	Parameter ini digunakan untuk membatasi kedalaman pohon keputusan.
<i>min_samples_leaf</i>	Untuk mengontrol jumlah minimum contoh data yang harus ada pada setiap daun (leaf) dari pohon keputusan.
<i>min_samples_split</i>	Untuk mengontrol jumlah minimum contoh data yang diperlukan untuk membagi simpul (node) dalam pohon keputusan.

2.2.12. Confusion Matrix

Confusion matrix merupakan salah satu metode yang digunakan untuk mengukur dan mengevaluasi performa kinerja dari metode klasifikasi

(*supervised learning*) pada *machine learning* dimana keluarannya dapat berupa dua kelas atau lebih (Karsito, 2019). *Confusion matrix* berbentuk tabel matriks yang menggambarkan kinerja dari suatu model klasifikasi pada rangkaian data uji yang menyatakan data uji yang dinyatakan terklasifikasi dengan benar dan data uji yang dinyatakan terklasifikasi salah. Terdapat empat istilah dalam merepresentasikan hasil proses klasifikasi pada *confusion matrix* yaitu sebagai berikut.

- *True Positive* (TP) : adalah jumlah data dengan nilai hasil positif dan nilai hasil prediksi positif.
- *False Positive* (FP) : adalah jumlah data dengan nilai hasil negatif dan nilai hasil prediksi positif.
- *True Negative* (TN) : adalah jumlah data dengan nilai hasil negatif dan nilai hasil prediksi negatif.
- *False Negative* (FN) : adalah jumlah data dengan nilai hasil positif dan nilai hasil prediksi negative

Tabel 5 *Confusion Matrix*

Fakta	Prediksi	
	Negatif	Positif
Negatif	TN (<i>True Negative</i>)	FP (<i>False Positive</i>)
positif	FN (<i>False Negative</i>)	TP (<i>True Positive</i>)

Representasi nilai dari *True Positive* dan *True Negative* akan memberikan hasil prediksi data yang bernilai benar, sedangkan untuk nilai dari *False Positive* dan *False Negative* akan memberikan hasil prediksi data yang bernilai salah. Positif dan negative merupakan suatu hasil prediksi dari model. (Nurmasani & Pristyanto, 2021) Nilai dari *True Positive*, *True Negative*, *False Positive* dan *False Negative* adalah nilai yang digunakan untuk menghitung evaluasi *performance metrics* yang berguna untuk

mengukur kinerja model yang telah dibuat. Terdapat beberapa *performance metrics* yang biasa digunakan yaitu sebagai berikut:

a. *Accuracy*

Accuracy adalah jumlah presentasi yang menggambarkan seberapa akurat model dapat mengklasifikasikan data dengan benar dari keseluruhan data. Nilai *accuracy* ini merupakan nilai yang mengolah data yang terprediksi dengan tepat ditambah dengan data yang terprediksi tidak tepat dibagi dengan jumlah keseluruhan data yang ada (Nugroho, 2019). Untuk menghitung tingkat akurasi dapat digunakan rumus dalam persamaan sebagai berikut.

$$\text{Akurasi (accuracy)} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

b. *Precision (Positive Predictive Value)*

Precision dapat digunakan untuk mengukur kinerja sistem dengan menghitung data yang terklasifikasi dengan tepat dan data yang ditemukan terklasifikasi tidak tepat. Data yang terklasifikasi dengan tepat dijadikan data acuan, kemudian dibagi dengan hasil prediksi False Positive atau data yang terprediksi tidak tepat maka akan diketahui kecocokan antara data acuan dengan data prediksi. Sehingga semakin banyak data acuan yang terprediksi tidak tepat ditemukan pada proses klasifikasi maka nilai dari *precision* akan semakin kecil (Nugroho, 2019). Untuk menghitung nilai *precision* dapat digunakan rumus dalam persamaan sebagai berikut.

$$\text{Presisi (precision)} = \frac{TP}{TP + FP} \times 100\%$$

c. *Recall atau Sensitivity*

Recall merupakan model untuk mengukur evaluasi kinerja sistem dalam menemukan sebuah informasi kembali. Dengan demikian rasio prediksi data yang terklasifikasi benar akan dibandingkan dengan keseluruhan data yang benar positif (Nugroho, 2019).

Semakin banyak data yang tepat tidak ditemukan maka nilai recall akan semakin kecil. Untuk menghitung nilai recall dapat digunakan rumus dalam persamaan sebagai berikut.

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

d. *F1-Score*

F1-Score merupakan perbandingan rata-rata precision dan recall. Untuk menghitung nilai formula f1-score dapat digunakan rumus dalam persamaan sebagai berikut (Saputro & Sari, 2019).

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

III. METODOLOGI PENELITIAN

3.1. Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan pada November 2022 sampai dengan Juni 2023 yang bertempat di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2. Dataset dan Perangkat Penelitian

3.2.1. Dataset

Data penelitian yang digunakan berupa data dari *Heart Disease Dataset* dan Penyakit Jantung. *Dataset* yang bersumber dari *Public Health Dataset*. *Dataset* yang digunakan dapat diakses melalui <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>.

Dataset ini merupakan tabular data yang terdiri dari 14 atribut. Total data pada dataset ini totalnya adalah 1025 data.

Berikut deskripsi mengenai atribut penelitian yang digunakan terdapat pada Tabel 5.

Tabel 6 Deskripsi Atribut Penelitian

No	Fitur	Jenis	Keterangan, Nilai
1	<i>Age</i>	Kontinu	Pasien usia , 28 ke 77
2	<i>Sex</i>	Diskrit	1 = Laki-laki; 0 = Perempuan
3	<i>CP (Chest Pain)</i>	Diskrit	Nilai 1: angina tipikal; Nilai 2: angina atipikal; Nilai 3: nyeri non angina; Nilai 4: tanpa gejala

No	Fitur	Jenis	Keterangan, Nilai
4	FBS (<i>Fasting Blood Sugar</i>)	Diskrit	Puasa Darah Gula, 0 = salah (FBS <120 mg/dl) 1 = benar (FBS > 120 mg/dl)
5	Restecg (<i>Result ECG</i>)	Diskrit	Nilai 0: normal; Nilai 1: memiliki gelombang ST-T kelainan (gelombang T inversi dan/atau ST elevasi atau depresi dari > 0,05 mV); Nilai2 : menunjukkan mungkin atau hipertrofi ventrikel kiri yang pasti oleh Kriteria Estes
6	Exang (<i>Exercise Induced Angina</i>)	Diskrit	Latihan diinduksi angina. 1 = Ya; 0 = Tidak
7	<i>Slope</i>	Diskrit	Ukuran kemiringan latihan puncak, Nilai 1: menanjak, Nilai 2: datar, Nilai 3: miring ke bawah
8	Ca	Diskrit	nomor kapal utama berwarna oleh flourosopy (0-3)
9	Thal	Diskrit	Kecepatan jantung pasien, 3 = biasa; 6=cacat tetap; 7 = dapat dibalik cacat
10	Trestbps	Kontinu	Tekanan darah istirahat pasien diukur dalam mm Hg saat masuk ke itu RSUD 80-200.
11	Chol	Kontinu	Pasien serum kolesterol diukur di mg/dl. 85, 100 - 200 - 394, 400 - 603
12	Thalach	Kontinu	Kecepatan maksimum jantung pasien tercapai.60-202, Rendah: di bawah 50, Biasa:51-119, Tinggi: 120-202 [6]
13	Oldpeak	Kontinu	ST depresi dibuat oleh olahraga relatif ke istirahat -2.6 ke -0,1, 0, 0,1 ke 6.2, 120
14	Diagnosis	Diskrit	0: Tidak ada Penyakit jantung; 1: Ada Penyakit Jantung

3.2.2. Perangkat Penelitian

Untuk melakukan penelitian ini, dibutuhkan *hardware* dan *software* sebagai berikut :

A. Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan dalam penelitian ini adalah sebuah laptop dengan spesifikasi Processor Intel Core i3-4030U CPU @ 1.90GHz, *Installed physical Memory* (RAM) 4GB, *system type x64-based PC*.

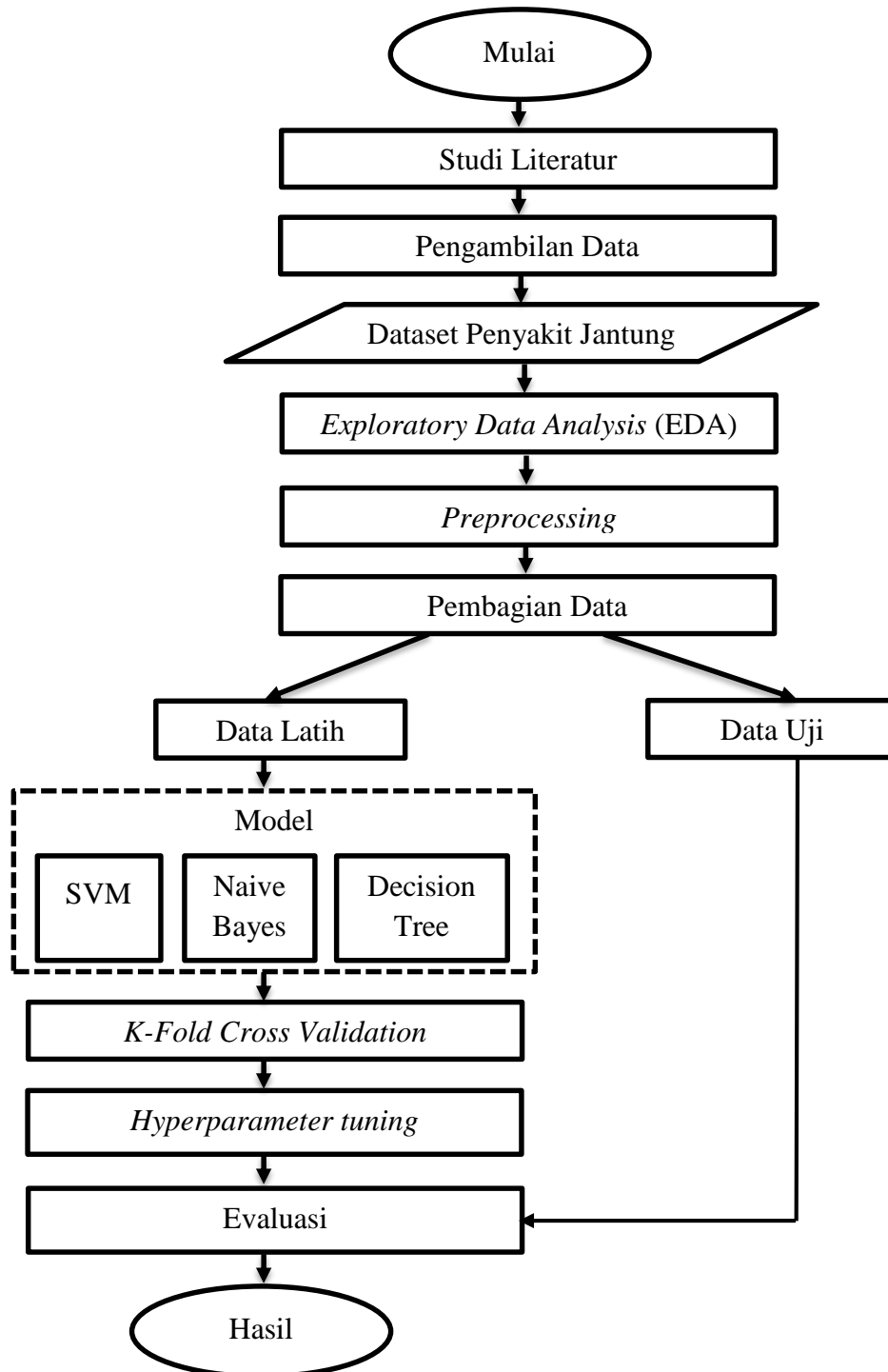
B. Perangkat Lunak (*Software*)

Adapun *software* yang digunakan pada penelitian ini, yakni:

- Sistem operasi *Windows 10*
- Microsoft Excel 2010 untuk menampilkan data.
- *Google Colab* dan *Jupyter Notebook* sebagai *tools* untuk mengolah dan menganalisis data penelitian.
- *Library* yang digunakan antara lain:
 - Pandas
 - NumPy
 - SelectKBest, chi2
 - train_test_split, cross_val_score
 - StandardScaler, LabelEncoder, MinMaxScaler
 - DecisionTreeClassifier
 - confusion_matrix, accuracy_score
 - GaussianNB
 - svm
 - seaborn as sns
 - classification_report
 - GridSearchCV

3.3. Alur Kerja Penelitian

Pada penelitian ini melalui beberapa tahapan sebagai berikut:



Gambar 3. Alur Penelitian

3.4. Studi Literatur

Studi Literatur adalah penelitian yang dilakukan oleh peneliti dengan mengumpulkan berbagai buku dan jurnal yang berkaitan dengan tujuan dan masalah penelitian. Langkah ini dimaksudkan sebagai referensi untuk memperkuat penelitian yang sedang berlangsung yang ada pada penelitian sebelumnya.

3.5. Pengumpulan Data

Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang digunakan adalah metode pengumpulan data sekunder. Data penelitian yang digunakan berupa data dari *Heart Disease Dataset*. *Dataset* yang bersumber dari *Public Health Dataset*. *Dataset* yang digunakan dapat diakses melalui.

<https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>.

3.6. Preprocessing

Tahapan selanjutnya adalah *preprocessing*. Pada tahap ini data yang masih berupa nilai numerik dan juga continue ditransformasikan ke dalam bentuk kategorik dan dibuat skala atau interval sehingga menghasilkan range yang lebih kecil sebagai bahan pembelajaran algoritma SVM, Naive Bayes dan *Decision Tree* akan lebih mudah diklasifikasikan dengan menggunakan Python sehingga memudahkan dalam memproses klasifikasi pada dataset yang telah tersedia sebelumnya.

Tahap pertama pada proses ini yaitu melakukan persiapan data. Dimana dalam persiapan data, data yang akan digunakan dilakukan pembagian variabel menjadi variabel dependen dan variabel independen. Dilanjutkan dengan normalisasi data agar data rentang antara 0 dan 1 untuk memudahkan perhitungan. Dan selanjutnya dilakukan pembagian data menjadi data training dan data testing.

Pada tahap ini, data training sebesar 80% dan data testing 20% dari dataset. Setelah tahapan pembagian data menjadi data training dan data testing, selanjutnya yaitu akan dilakukan proses transpose matrix.

3.7. Penerapan Algoritma SVM, Naive Bayes dan *Decision Tree*

Pada tahap ini memilih dan menerapkan teknik yang tepat untuk mendapatkan hasil yang optimal, pada penelitian ini algoritma yang digunakan adalah algoritma SVM, Naive Bayes dan *Decision Tree* dilakukan dengan beberapa percobaan dengan menggunakan operator split data dan tahap ini dilakukan dengan menggunakan Google Colab dan Jupyter Notebook sebagai tools dan python sebagai bahasa pemrograman.

1. *Support Vector Machine* (SVM)

Pada tahap ini, hal yang dilakukan terlebih dahulu yaitu menentukan tipe *kernel* (*linear*, *polynomial* dan *Gaussian*) serta menentukan nilai parameter. Selanjutnya *training* SVM mulai dilakukan pada data model SVM klasifikasi. Jika model terbaik pada data telah terpilih maka sudah bisa untuk dilakukan proses klasifikasi.

2. Naive Bayes

Setelah tahap *preprocessing*, maka tahapan berikutnya dalam proses perhitungan Naive Bayes adalah menghitung *Joint Probability Distribution* suatu gejala untuk dapat mengetahui apakah pasien tersebut didiagnosis penyakit jantung atau tidak. Pertama-tama masukkan perhitungan peluang dan prior sesuai dengan parameter yang diinput data baru tersebut mulai dari peluang penyakit jantung iya dan tidak lalu perhitungan prior semua variabel yang ada.,

Peluang Diagnosis Ya :

$$P(D | Y_a) = \frac{P(\text{Jumlah klasifikasi Ya pada Preprocessing})}{\text{Jumlah data}}$$

Prior Age :

$$P(\text{Dewasa} \mid Ya) = \frac{P(\text{Jumlah data usia Dewasa YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Sex :

$$P(\text{Perempuan} \mid Ya) = \frac{P(\text{Jumlah data jenis kelamin perempuan YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Cp :

$$P(\text{Asymptomatic Pain} \mid Ya) = \frac{P(\text{Jumlah data CP Asymptomatic YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Trestbps :

$$P(\text{Normal} \mid Ya) = \frac{P(\text{Jumlah data Trestbps Normal YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Chol :

$$P(\text{Normal} \mid Ya) = \frac{P(\text{Jumlah data Chol Normal YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Fbs :

$$P(\text{Normal} \mid Ya) = \frac{P(\text{Jumlah data Fbs Normal YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Restecg :

$$P(\text{Normal} \mid Ya) = \frac{P(\text{Jumlah data Restecg Normal YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Thalach :

$$P(\text{Tinggi} \mid Ya) = \frac{P(\text{Jumlah data Thalach tinggi YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Exang :

$$P(\text{Tidak nyeri} \mid Ya) = \frac{P(\text{Jumlah data exang tidak nyeri YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Oldpeak :

$$P(\text{Normal} \mid Ya) = \frac{P(\text{Jumlah data Oldpeak Normal YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Slope :

$$P(\text{Up} \mid Ya) = \frac{P(\text{Jumlah data Slope Up YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Ca :

$$P(0 \mid Ya) = \frac{P(\text{Jumlah data Ca 0 YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Prior Thall :

$$P(\text{Normal} \mid Ya) = \frac{P(\text{Jumlah data Thall Normal YA D})}{\text{Jumlah data klasifikasi YA D}}$$

Peluang Diagnosis Tidak :

$$P(D | \text{Tidak}) = \frac{P(\text{Jumlah klasifikasi TIDAK pada Preprocessing})}{\text{Jumlah data}}$$

Prior Age :

$$P(\text{Dewasa} | \text{Tidak}) = \frac{P(\text{Jumlah data usia Dewasa TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Sex :

$$P(\text{Perempuan} | \text{Tidak}) = \frac{P(\text{Jumlah data jenis kelamin perempuan TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Cp :

$$P(\text{Asymptomatic Pain} | \text{Tidak}) = \frac{P(\text{Jumlah data CP Asymptomatic TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Trestbps :

$$P(\text{Normal} | \text{Tidak}) = \frac{P(\text{Jumlah data Trestbps Normal YA D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Chol :

$$P(\text{Normal} | \text{Tidak}) = \frac{P(\text{Jumlah data Chol Normal TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Fbs :

$$P(\text{Normal} | \text{Tidak}) = \frac{P(\text{Jumlah data Fbs Normal TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Restecg :

$$P(\text{Normal} | \text{Tidak}) = \frac{P(\text{Jumlah data Restecg Normal TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Thalach :

$$P(\text{Tinggi} | \text{Tidak}) = \frac{P(\text{Jumlah data Thalach tinggi TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Exang :

$$P(\text{Tidak nyeri} | \text{Tidak}) = \frac{P(\text{Jumlah data exang tidak nyeri TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Oldpeak :

$$P(\text{Normal} | \text{Tidak}) = \frac{P(\text{Jumlah data Oldpeak Normal TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Slope :

$$P(\text{Up} | \text{Tidak}) = \frac{P(\text{Jumlah data Slope Up TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Ca :

$$P(0 | \text{Tidak}) = \frac{P(\text{Jumlah data Ca 0 TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

Prior Thal :

$$P(\text{Normal} \mid \text{Tidak}) = \frac{P(\text{Jumlah data Thal Normal TIDAK D})}{\text{Jumlah data klasifikasi TIDAK D}}$$

3. *Decision Tree*

Algoritma C4.5 untuk menciptakan pohon keputusan sebagai berikut yaitu:

- a. Pemilihan atribut yang akan dijadikan sebagai akar
- b. Pembuatan cabang untuk setiap nilai
- c. Pembagian kasus dalam cabang
- d. Repetisi proses sampai setiap cabang guna agar semua kasus pada cabang memiliki kelas yang sama.

3.8. Hasil

Hasil dari penelitian ini dengan menggunakan penerapan algoritma *Support Vector Machine* (SVM), Naive Bayes dan *Decision Tree* untuk mengetahui algoritma mana yang memiliki tingkat akurasi yang lebih tinggi. Pengukuran dilakukan menggunakan metode *Confusion Matrix*.

V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Pada perbandingan *Support Vector Machine*, Naive Bayes dan *Decision Tree* dalam mengkategorikan penyakit jantung, adalah sebagai berikut:

1. Model pada prediksi dengan klasifikasi penyakit jantung diterapkan dengan menggunakan metode *Support Vector Machine* (SVM), Naive Bayes dan *Decision Tree*.
2. Pada penelitian prediksi untuk klasifikasi untuk diagnosa penyakit jantung, metode *Support Vector Machine* (SVM) dan *Decision Tree* memiliki performa ketepatan prediksi sangat baik engan memperoleh nilai akurasi sebesar 99% dan metode Naive Bayes yang memperoleh nilai akurasi yaitu 84%.
3. Pada penelitian ini, digunakan *Hyperparameter tuning* untuk membantu meningkatkan kinerja model dengan menemukan kombinasi parameter terbaik yang sesuai dengan data yang digunakan. Dengan mengoptimalkan *hyperparameter*, model dapat mencapai tingkat akurasi, presisi, recall, dan F1-score yang lebih tinggi.

5.2. Saran

Adapun saran yang diberikan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini dapat dikembangkan lagi dengan berbagai macam metode klasifikasi lainnya, dengan harapan dapat menemukan hasil klasifikasi yang lebih baik sebagai bahan perbandingan.
2. Pentingnya mengetahui struktur dataset yang digunakan dalam penelitian karenadata tersebut dapat mengubah keakuratan hasil.

DAFTAR PUSTAKA

- Adi, S. (2022). Komparasi Metode Support Vector Machine (Svm), K-Nearest Neighbors (Knn), Dan Random Forest (Rf) Untuk Prediksi Penyakit Gagal Jantung. *Jurnal Ilmiah Matematika*, 10(02), 258–268.
- Afrizal, M. R., Nugroho, R. A., Kartini, D., Herteno, R., & Selatan, K. (1979). Xgboost Dengan Random Search Hyper- Parameter Tuning Untuk Klasifikasi Situs. 15, 40–47.
- Akbar, J. A., Fadela, Z. I., Fachruddin, L., Ardiansyah, F., & Bakhri, Z. M. (2019). Analisis Penyakit Jantung Koroner Menggunakan Decision Tree. *Departemen Ilmu Komputer, IPB University, Cvd*, 3–7.
- Ali, L., Khan, S. U., Golilarz, N. A., Yakubu, I., Qasim, I., Noor, A., & Nour, R. (2019). A Feature-Driven Decision Support System for Heart Failure Prediction Based on χ^2 Statistical Model and Gaussian Naive Bayes. *Computational and Mathematical Methods in Medicine*, 2019. <https://doi.org/10.1155/2019/6314328>
- Al-Janabi, M. I., Qutqut, M. H., & Hijjawi, M. (2018). Machine learning classification techniques for heart disease prediction: a review. *International Journal of Engineering & Technology*, 7(4), 5373–5379. <https://doi.org/10.14419/ijet.v7i4.28646>
- Anderies, A., Tchin, J. A. R. W., Putro, P. H., Darmawan, Y. P., & Gunawan, A. A. S. (2022). Prediction of Heart Disease UCI Dataset Using Machine Learning Algorithms. *Engineering, Mathematics and Computer Science (EMACS) Journal*, 4(3), 87–93. <https://doi.org/10.21512/emacsjournal.v4i3.8683>

- Arifin, O., & Sasongko, T. B. (2018). Analisa perbandingan tingkat performansi metode support vector machine dan naïve bayes classifier. *Seminar Nasional Teknologi Informasi Dan Multimedia 2018*, 6(1), 67–72.
- Chandra Reddy, N. S., Shue Nee, S., Zhi Min, L., & Xin Ying, C. (2019). Classification and Feature Selection Approaches by Machine Learning Techniques: Heart Disease Prediction. *International Journal of Innovative Computing*, 9(1), 39–46. <https://doi.org/10.11113/ijic.v9n1.210>
- Dona, D., Maradona, H., & Masdewi, M. (2021). Sistem Pakar Diagnosa Penyakit Jantung Dengan Metode Case Based Reasoning (Cbr). *ZONAsi: Jurnal Sistem Informasi*, 3(1), 1–12. <https://doi.org/10.31849/zn.v3i1.6442>
- Golpour, P., Ghayour-Mobarhan, M., Saki, A., Esmaily, H., Taghipour, A., Tajfard, M., Ghazizadeh, H., Moohebaty, M., & Ferns, G. A. (2020). Comparison of support vector machine, naïve bayes and logistic regression for assessing the necessity for coronary angiography. *International Journal of Environmental Research and Public Health*, 17(18), 1–9. <https://doi.org/10.3390/ijerph17186449>
- Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., Sun, R., & Garcíá-Magarinõ, I. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, 2018. <https://doi.org/10.1155/2018/3860146>
- Jeyaganesan, J., Sathiya, A., Keerthana, S., & Aiyer, A. (2020). *Diagnosis And Prediction Of Heart Disease Using Machine Learning Techniques*. 19(2), 1817–1827. <https://doi.org/10.17051/ilkonline.2020.02.696765>
- Kaggle.com, 22 September 2022, <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>
- Khennou, F., Fahim, C., Chaoui, H., & Chaoui, N. E. H. (2019). A machine learning approach: Using predictive analytics to identify and analyze high risks patients with heart disease. *International Journal of Machine Learning and Computing*, 9(6), 762–767. <https://doi.org/10.18178/ijmlc.2019.9.6.870>

- Maulidah, N., Supriyadi, R., Utami, D. Y., Hasan, F. N., Fauzi, A., & Christian, A. (2021). Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes. *Indonesian Journal on Software Engineering (IJSE)*, 7(1), 63–68. <https://doi.org/10.31294/ijse.v7i1.10279>
- Muis, I. A., Affandes, M., Muis, I. A., Studi, P., Informatika, T., & Sains, F. (2015). Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet. 12(2), 189–197.
- Naryadi, N. W. J. (2019). Hubungan Tingkat Pengetahuan, Tingkat Dukungan Keluarga Dan Tingkat Kepatuhan Diet Pasien Jantung Pasca Rawat Inap Di Rumah Sakit Umum Bangli. 2016, 67. <http://repository.poltekkes-denpasar.ac.id/id/eprint/3142>
- Nugraha, W., & Sasongko, A. (2022). Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search. *SISTEMASI: Jurnal Sistem Informasi*, 11(2), 391–401.
- Nurmasani, A., & Pristyanto, Y. (2021). Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class. *Pseudocode*, 8(1), 21–26. <https://doi.org/10.33369/pseudocode.8.1.21-26>
- Pisner, D. A., & Schnyer, D. M. (2019). Support vector machine. *Machine Learning: Methods and Applications to Brain Disorders*, 101–121. <https://doi.org/10.1016/B978-0-12-815739-8.00006-7>
- Putri, T. A. E., Widiharih, T., & Santoso, R. (2023). Penerapan Tuning Hyperparameter Randomsearchcv Pada Adaptive Boosting Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung. *Jurnal Gaussian*, 11(3), 397–406. <https://doi.org/10.14710/j.gauss.11.3.397-406>
- Radhi, M., Amalia, A., Sitompul, D. R. H., Sinurat, S. H., & Indra, E. (2022). Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 4(2), 23–27. <https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2475>

- Sari, E. D. N., & Irhamah, I. (2020). Analisis Sentimen Nasabah pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, Naïve Bayes Classifier (NBC), dan Support Vector Machine (SVM). *Jurnal Sains Dan Seni ITS*, 8(2). <https://doi.org/10.12962/j23373520.v8i2.44565>
- Science, C. (2019). *Prediksi Penyakit Jantung dengan Algoritma Klasifikasi*. 5(1), 978–979.
- Supriyadi, R., Maulidah, N., Fauzi, A., Nalatissifa, H., & Diantika, S. (2022). 55~59 Diterima Februari 1. *JURNAL SWABUMI*, 10(1), 2022. <https://archive.ics.uci.edu/ml/datasets/Autis>
- Wibisono, A. B., & Fahrurozi, A. (2019). Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 24(3), 161–170. <https://doi.org/10.35760/tr.2019.v24i3.2393>
- Yadav, D. C., & Pal, S. (2021). Analysis of Heart Disease Using Parallel and Sequential Ensemble Methods With Feature Selection Techniques. *International Journal of Big Data and Analytics in Healthcare*, 6(1), 40–56. <https://doi.org/10.4018/ijbdah.20210101.oa4>