

**PREDIKSI PENYAKIT JANTUNG KORONER DENGAN METODE
NAIVE BAYES, SVM DAN RANDOM FOREST**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi *SI-Informatika*



disusun oleh

MUHAMMAD IQBAL PRATAMA

18.11.2283

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**PREDIKSI PENYAKIT JANTUNG KORONER DENGAN METODE
NAIVE BAYES, SVM DAN RANDOM FOREST
SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi *SI- Informatika*



disusun oleh

MUHAMMAD IQBAL PRATAMA

18.11.2283

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**PREDIKSI PENYAKIT JANTUNG KORONER DENGAN METODE NAIVE BAYES,
SVM DAN RANDOM FOREST**

yang disusun dan diajukan oleh

Muhammad Iqbal Pratama

18.11.2283

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Juli 2024

Dosen Pembimbing,



Ike Verawati, M.Kom
NIK. 190302237

HALAMAN PENGESAHAN
SKRIPSI
PREDIKSI PENYAKIT JANTUNG KORONER DENGAN METODE
NAIVE BAYES, SVM DAN RANDOM FOREST

yang disusun dan diajukan oleh

Muhammad Iqbal Pratama

18.11.2283

Telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Juli 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Melany Mustika Dewi, M.Kom
NIK. 190302455



Anggit Ferdita Nugraha, S.T., M.Eng
NIK. 190302480



Ike Verawati, M.Kom
NIK. 190302237



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Juli 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhammad Iqbal Pratama
NIM : 18.11.2283

Menyatakan bahwa Skripsi dengan judul berikut:

PREDIKSI PENYAKIT JANTUNG DENGAN METODE NAIVE BAYES, SVM DAN RANDOM FOREST

Dosen Pembimbing : Ike Verawati, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 26 Juli 2024

Yang Menyatakan,



Muhammad Iqbal Pratama

HALAMAN PERSEMBAHAN

Puji syukur kepada Tuhan Yang Maha Esa atas berkah, rahmat serta karunia-Nya yang telah memberikan kemudahan sehingga saya di dapat menyelesaikan skripsi ini sehingga dapat diselesaikan dengan baik. Dengan ini saya persembahkan skripsi ini kepada semua pihak yang turut mendukung perkuliahan hingga mampu menyelesaikan studi untuk meraih gelar sarjana yaitu:

1. Kedua Orang tua yang telah memberikan doa, menyemangati dan memberikan dukungan setiap waktunya.
2. Keluarga besar yang selalu menyemangati penulis.
3. Teman – teman Kelas Informatika 07 terutama teman teman yang pernah menjadi tim kelompok dalam mengerjakan tugas, dan membantu kelancaran kuliah.
4. Serta semua teman dekat penulis yang tidak dapat disebutkan satu-persatu telah memberikan dukungan yang tidak ada hentinya dan selalu mendampingi menyemangati selalu.
5. Untuk semua pihak maupun team yang tidak bisa penulis sebutkan satu satu.

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa, atas segala rahmat, hidayah, dan karunia-Nya yang senantiasa melimpahkan berkah dalam setiap langkah perjalanan hidup kami. Shalawat serta salam semoga tercurah kepada baginda Nabi Muhammad SAW, insan yang menjadi tauladan dalam kehidupan ini.

Penelitian ini merupakan upaya kami untuk menggali lebih dalam dalam bidang prediksi penyakit jantung, suatu hal yang menjadi fokus perhatian dalam dunia kedokteran dan kesehatan masyarakat. Penyakit jantung merupakan salah satu penyakit yang menjadi penyebab utama kematian di seluruh dunia. Dengan adanya kemajuan teknologi dan metode analisis data, kami tertarik untuk mengeksplorasi potensi metode-metode seperti Naive Bayes, Random Forest, dan Support Vector Machine (SVM) dalam memprediksi kemungkinan penyakit jantung pada individu.

Dalam penyusunan penelitian ini, kami berusaha untuk menggabungkan keahlian dari berbagai disiplin ilmu, termasuk ilmu komputer, statistika, dan kedokteran. Melalui kolaborasi ini, kami berharap dapat memberikan kontribusi yang bermanfaat bagi perkembangan ilmu pengetahuan dan kesejahteraan umat manusia.

Kami menyadari bahwa penelitian ini masih jauh dari kesempurnaan. Oleh karena itu, segala kritik dan saran yang membangun sangat kami harapkan untuk perbaikan ke depannya. Semoga penelitian ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dan kesejahteraan umat manusia.

Akhir kata, kami berdoa semoga penelitian ini dapat bermanfaat bagi banyak pihak. Semoga Allah SWT senantiasa memberikan rahmat dan hidayah-Nya kepada kita semua. Amin.

Yogyakarta, 26 Juli 2024

Penulis

Muhammad Iqbal Pratama

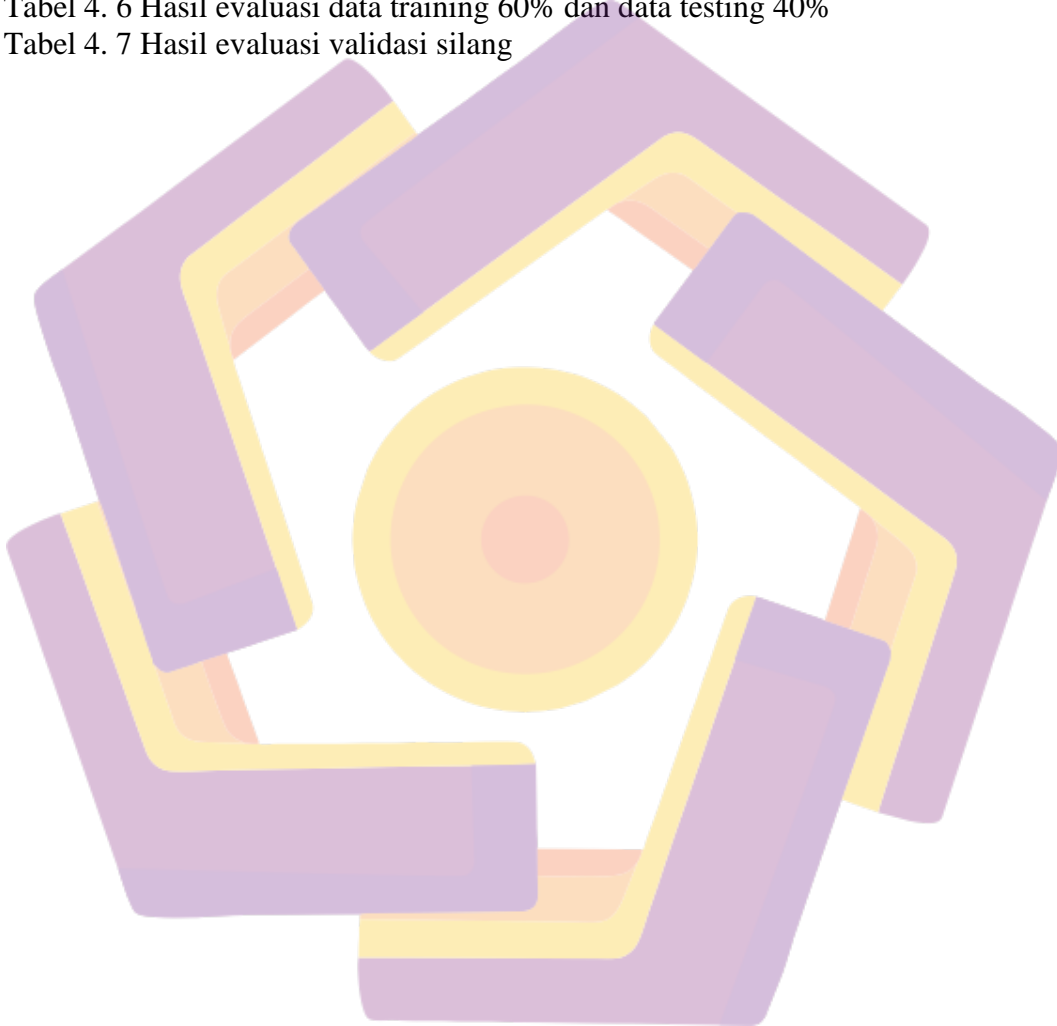
DAFTAR ISI

HALAMAN JUDUL	1
HALAMAN PERSETUJUAN	2
HALAMAN PENGESAHAN	3
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	4
HALAMAN PERSEMBAHAN	5
KATA PENGANTAR	6
DAFTAR ISI	8
DAFTAR TABEL	10
DAFTAR GAMBAR	11
DAFTAR LAMPIRAN	13
DAFTAR LAMBANG DAN SINGKATAN	14
INTISARI	15
ABSTRACT	16
BAB I	
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II	
TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	12
2.2.1 Penyakit Jantung Koroner	12
2.2.2 Klasifikasi	12
2.2.3 Prediksi	13
2.2.4 Algoritma	13
2.2.5 Machine Learning	13
2.2.6 Confusion Matrix	15
2.2.7 Naive Bayes	16
2.2.8 Random Forest	17
2.2.9 Support Vector Machine	18
2.2.10 Python	19
BAB III	
METODE PENELITIAN	20
3.1 Alur Penelitian	20
3.2 Alat dan Bahan	22

3.3.1 Perangkat Keras	22
3.3.2 Perangkat Lunak	22
3.3.3 Bahan	23
BAB IV	
HASIL DAN PEMBAHASAN	26
4.1 Pengumpulan data	26
4.2 Preprocessing	27
4.3 Pembagian data	28
4.4 Pembuatan model algoritma	30
4.4.1 Perancangan model algoritma naive bayes	30
4.4.2 Perancangan model algoritma random forest	33
4.4.3 Perancangan model algoritma support vector machine	36
4.5 Hasil perancangan model algoritma	39
4.6 Hasil Evaluasi	53
BAB V	
PENUTUP	58
5.1 Kesimpulan	58
5.2 Saran	60
REFERENSI	61
LAMPIRAN	63

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	7
Tabel 2. 2 Confusion Matrix Table	12
Tabel 3. 1 Perangkat Keras	22
Tabel 3. 2 Deskripsi dataset	23
Tabel 4. 1 Data latih 80% dan data uji 20%	33
Tabel 4. 2 Data latih 70% dan data uji 30%	33
Tabel 4. 3 Data latih 60% dan data uji 40%	34
Tabel 4. 4 Hasil evaluasi data training 80% dan data testing 20%	86
Tabel 4. 5 Hasil evaluasi data training 70% dan data testing 30%	87
Tabel 4. 6 Hasil evaluasi data training 60% dan data testing 40%	89
Tabel 4. 7 Hasil evaluasi validasi silang	90



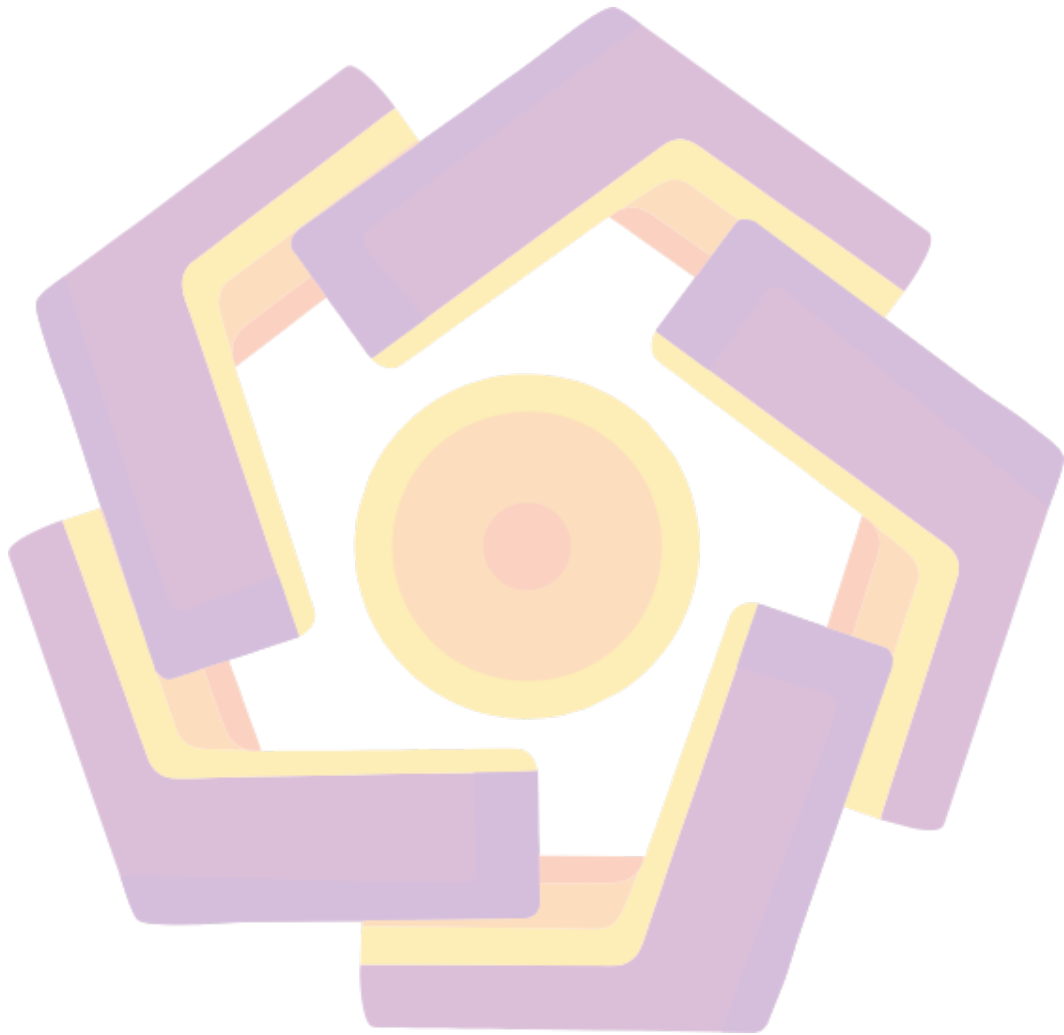
DAFTAR GAMBAR

Gambar 2. 1 Random Forest.	15
Gambar 3. 1 Alur Penelitian.	20
Gambar 3. 2 Dataset.	23
Gambar 3. 3 Statistik deskriptif.	25
Gambar 3. 4 Distribusi fitur numerik.	26
Gambar 3. 5 Distribusi kelas target	27
Gambar 3. 6 Keseimbangan fitur berdasarkan target	28
Gambar 3. 7 Heatmap korelasi fitur	29
Gambar 4. 1 Dataset.	30
Gambar 4. 2 Pembersihan data.	31
Gambar 4. 3 Seleksi data.	31
Gambar 4. 4 Transformasi data.	32
Gambar 4. 5 Data train dan data test (80:20).	33
Gambar 4. 6 Data train dan data test (70:30).	34
Gambar 4. 7 Data train dan data test (60:40).	34
Gambar 4. 8 Import library model naive bayes.	35
Gambar 4. 9 Sinkronisasi akun google.	35
Gambar 4. 10 import dataset ke dalam google collab.	36
Gambar 4. 11 fitur scaling.	36
Gambar 4. 12 Model naive bayes.	36
Gambar 4. 13 Print klasifikasi report.	37
Gambar 4. 14 Membuat Visualisasi confusion matrix.	37
Gambar 4. 15 Import library random forest.	37
Gambar 4. 16 Sinkronisasi akun google.	38
Gambar 4. 17 import dataset ke dalam google collab.	38
Gambar 4. 18 fitur scaling.	38
Gambar 4. 19 model random forest.	39
Gambar 4. 20 Print klasifikasi report.	39
Gambar 4. 21 Membuat Visualisasi confusion matrix.	39
Gambar 4. 22 import library svm.	40
Gambar 4. 23 Sinkronisasi akun google.	40
Gambar 4. 24 import dataset ke dalam google collab.	40
Gambar 4. 25 fitur scaling.	41
Gambar 4. 26 model svm.	41
Gambar 4. 27 Print klasifikasi report.	41
Gambar 4. 28 Membuat Visualisasi confusion matrix.	41
Gambar 4. 29 Hasil model naive bayes 80:20.	42
Gambar 4. 30 Hasil confusion matrix naive bayes 80:20.	43
Gambar 4. 31 Hasil model random forest 80:20.	44
Gambar 4. 32 Hasil confusion matrix random forest 80:20.	45
Gambar 4. 33 Hasil model svm 80:20.	46
Gambar 4. 34 Hasil confusion matrix svm 80:20.	47
Gambar 4. 35 Hasil model logistic regression 80:20.	48
Gambar 4. 36 Hasil confusion matrix logistic regression 80:20.	49
Gambar 4. 37 Hasil model decision tree 80:20.	50
Gambar 4. 38 Hasil confusion matrix decision tree 80:20.	50
Gambar 4. 39 Hasil model gradient boosting 80:20.	51
Gambar 4. 40 Hasil confusion matrix svm 80:20.	52

Gambar 4. 41 Hasil model naive bayes 70:30.	53
Gambar 4. 42 Hasil confusion matrix naive bayes 70:30.	54
Gambar 4. 43 Hasil model random forest 70:30.	55
Gambar 4. 44 Hasil confusion matrix random forest 70:30.	56
Gambar 4. 45 Hasil model svm 70:30.	57
Gambar 4. 46 Hasil confusion matrix svm 70:30.	58
Gambar 4. 47 Hasil model logistic regression 70:30.	59
Gambar 4. 48 Hasil confusion matrix logistic regression 70:30.	59
Gambar 4. 49 Hasil model decision tree 70:30.	60
Gambar 4. 50 Hasil confusion matrix decision tree 70:30.	61
Gambar 4. 51 Hasil model gradient boosting 70:30.	62
Gambar 4. 52 Hasil confusion matrix gradient boosting 70:30.	62
Gambar 4. 53 Hasil model naive bayes 60:40.	63
Gambar 4. 54 Hasil confusion matrix naive bayes.	64
Gambar 4. 55 Hasil model random forest 60:40.	65
Gambar 4. 56 Hasil confusion matrix random forest.	66
Gambar 4. 57 Hasil model svm 60:40.	67
Gambar 4. 58 Hasil confusion matrix svm.	68
Gambar 4. 59 Hasil model logistic regression 60:40.	69
Gambar 4. 60 Hasil confusion matrix logistic regression 60:40.	70
Gambar 4. 61 Hasil model decision tree 60:40.	71
Gambar 4. 62 Hasil confusion matrix decision tree 60:40.	71
Gambar 4. 63 Hasil model gradient boosting 60:40.	72
Gambar 4. 64 Hasil confusion matrix gradient boosting 60:40.	73
Gambar 4. 65 Hasil cross validation score dan rata-rata akurasi naive bayes.	74
Gambar 4. 66 Hasil klasifikasi report validasi silang naive bayes.	74
Gambar 4. 67 Hasil confusion matrix validasi silang naive bayes.	75
Gambar 4. 68 Hasil cross validation dan rata-rata akurasi naive bayes.	76
Gambar 4. 69 Hasil klasifikasi report validasi silang random forest.	76
Gambar 4. 70 Hasil confusion matrix validasi silang random forest.	77
Gambar 4. 71 Hasil cross validation score dan rata-rata akurasi svm.	78
Gambar 4. 72 Hasil klasifikasi report validasi silang svm.	78
Gambar 4. 73 Hasil confusion matrix validasi silang svm	79
Gambar 4. 74 Hasil cross validation score dan rata-rata akurasi logistic regression.	80
Gambar 4. 75 Hasil klasifikasi report validasi silang logistic regression.	80
Gambar 4. 76 Hasil confusion matrix validasi silang logistic regression.	81
Gambar 4. 77 Hasil cross validation score dan rata-rata akurasi decision tree.	82
Gambar 4. 78 Hasil klasifikasi report validasi silang decision tree.	82
Gambar 4. 79 Hasil confusion matrix validasi silang decision tree.	83
Gambar 4. 80 Hasil cross validation score dan rata-rata akurasi gradient boosting.	84
Gambar 4. 81 Hasil klasifikasi report validasi silang gradient boosting.	84
Gambar 4. 82 Hasil confusion matrix validasi silang logistic regression.	85
Gambar 4. 83 Grafik diagram 80:20.	87
Gambar 4. 84 Grafik diagram 70:30.	88
Gambar 4. 85 Grafik diagram 60:40.	90
Gambar 4. 86 Grafik diagram validasi silang	91

DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
RF	Random Forest



INTISARI

Penyakit jantung koroner (PJK) merupakan salah satu penyebab utama kematian global. Studi ini bertujuan untuk membandingkan performa tiga algoritma klasifikasi dalam memprediksi PJK, yaitu Naive Bayes, Random Forest, dan SVM. Data yang digunakan berasal dari UCI machine learning repository dengan judul heart disease.

Metode penelitian menggunakan langkah-langkah standar dalam pemrosesan data dan pengujian model. Pertama, pembersihan data, seleksi data, dan transformasi data. Kemudian, dilakukan pembagian data menjadi 3 kelompok yaitu 80:20, 70:30, dan 60:40 untuk melatih dan menguji ketiga algoritma klasifikasi supaya mendapatkan hasil model yang maksimal.

Hasil penelitian menunjukkan bahwa ketiga algoritma klasifikasi memberikan performa yang berbeda-beda dalam memprediksi PJK. Naive bayes dan svm menunjukkan hasil yang sama pada pembagian data yaitu 80:20 dengan tingkat akurasi 89%, dan 60:40 dengan tingkat akurasi 84% sedangkan random forest mendapatkan tingkat akurasi 87%. Untuk pembagian data 70:30 model naive bayes lebih unggul dari random forest dan svm dengan tingkat akurasi 86%.

Penelitian ini memberikan wawasan tentang penggunaan berbagai algoritma klasifikasi untuk prediksi PJK, dengan implikasi potensial untuk pengembangan sistem pendukung keputusan di bidang medis. Namun, perlu diperhatikan bahwa performa algoritma dapat bervariasi tergantung pada karakteristik data spesifik dan parameter yang diatur. Penelitian selanjutnya dapat melibatkan dataset yang lebih besar atau faktor risiko tambahan untuk meningkatkan ketepatan prediksi dan generalisasi model.

Kata kunci: naive bayes, random forest, svm, prediksi, penyakit jantung koroner.

ABSTRACT

Coronary heart disease (CHD) is one of the main causes of death globally. This study aims to compare the performance of three classification algorithms in predicting CHD, namely Naive Bayes, Random Forest, and SVM. The data used comes from the UCI machine learning repository with the title heart disease.

The research method uses standard steps in data processing and model testing. First, data cleaning, data selection, and data transformation. Then, the data was divided into 3 groups, namely 80:20, 70:30, and 60:40 to train and test the three classification algorithms in order to get maximum model results.

The research results show that the three classification algorithms provide different performance in predicting CHD. Naive Bayes and SVM show the same results in data division, namely 80:20 with an accuracy rate of 89%, and 60:40 with an accuracy rate of 84%, while random forest gets an accuracy rate of 87%. For a 70:30 data division, the Naive Bayes model is superior to random forest and SVM with an accuracy rate of 86%.

This study provides insight into the use of various classification algorithms for CHD prediction, with potential implications for the development of decision support systems in the medical field. However, it should be noted that the performance of the algorithm may vary depending on the specific data characteristics and the parameters set. Future research could involve larger datasets or additional risk factors to improve prediction accuracy and generalization of the model.

Keyword: naive bayes, random forest, svm, prediction, coronary heart disease