

PREDIKSI RISIKO PENYAKIT JANTUNG MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

1st Ayu Febriana Lingga

Informatika

Universitas Multimedia Nusantara

Tangerang, Indonesia

ayu.febriana@student.umn.ac.id

2nd Nayasha Clarisa Dwisutrisna

Informatika

Universitas Multimedia Nusantara

Tangerang, Indonesia

nayasha.clarisa@student.umn.ac.id

3rd Salwa Putri Riswana

Informatika

Universitas Multimedia Nusantara

Tangerang, Indonesia

salwa.putri@student.umn.ac.id

4th Wilcoustine Qhrstmas Pniel Wijaya

Informatika

Universitas Multimedia Nusantara

Tangerang, Indonesia

wilcoustine@student.umn.ac.id

Abstract—Heart disease is a condition where there is a disturbance in the functioning of the heart. Heart disease itself has many types, such as cardiovascular, coronary heart, and sudden heart attacks. Heart disease is one of the illnesses that frequently occurs in the general population regardless of age, gender, or lifestyle. It can be caused by various factors, such as blockages in blood vessels, inflammation in the system, heart infections, or other abnormalities. According to the World Health Organization (WHO), which is a global health organization, heart disease is a major cause of death in countries like the United Kingdom, United States, Canada, and Australia. Currently, the number of adults diagnosed with heart disease has reached up to 26.6 million people, equivalent to 11.3% of the adult population. Therefore, early diagnosis of heart disease is crucial. However, manual prediction can be considered less effective due to a lack of medical staff expertise, which can lead to incorrect predictions. With the passage of time, advancing technology has greatly assisted the healthcare sector, especially in handling various diseases. In the predictive system for assessing the risk of heart disease developed by the author, the Support Vector Machine (SVM) method, which is part of machine learning, will be used. The author's dataset titled 'HEART DISEASE' comprises 14 features and covers 4 regions: Cleveland, Hungary, Switzerland, and Long Beach V..

Index Terms—Heart Disease, Machine Learning, SVM (Support Vector Machine), Dataset, Prediction

Abstrak—Penyakit jantung adalah sebuah situasi dimana terdapat gangguan terhadap fungsi kerja jantung. Penyakit jantung sendiri memiliki banyak jenis misalnya kardiovaskuler, jantung koroner dan serangan jantung secara tiba-tiba. Penyakit jantung menjadi salah satu penyakit yang kasusnya sering terjadi di khalayak umum tanpa memandang usia, jenis kelamin dan gaya hidup. Penyakit jantung sendiri dapat disebabkan oleh banyak hal, misalnya penyumbatan terhadap pembuluh darah, peradangan pada sistem, infeksi pada jantung, ataupun kelainan lainnya. Berdasarkan WHO (*World Health Organization*) yang merupakan sebuah organisasi kesehatan dunia menuliskan bahwa penyakit jantung adalah salah satu penyebab utama kematian di negara Inggris, Amerika Serikat, Kanada dan Australia. Hingga saat ini jumlah orang dewasa yang didiagnosis dengan penyakit jantung mencapai hingga 26,6 Juta Jiwa atau setara dengan

11,3% dari populasi orang dewasa. Dengan demikian, diagnosa dini terhadap penyakit jantung sangat penting untuk dilakukan. Namun, prediksi secara manual dapat dikatakan kurang efektif akibat kurangnya keahlian staff medis yang dapat menghasilkan prediksi yang salah. Seiring dengan berkembangnya waktu, teknologi yang semakin maju juga sangat membantu dalam bidang kesehatan terlebih lagi untuk menangani berbagai penyakit. Pada sistem prediksi risiko penyakit jantung yang disusun oleh penulis, akan digunakannya metode SVM (*Support Vector Machine*) yang menjadi bagian dari *machine learning*. Data yang digunakan oleh penulis merupakan dataset dengan judul "HEART DISEASE" yang memiliki 14 fitur dan mencakup 4 wilayah yaitu Cleveland, Hungaria, Switzerland, dan Long Beach V.

Kata Kunci—Heart Disease, Machine Learning, SVM (Support Vector Machine), Dataset, Prediction

I. PENDAHULUAN

Penyakit jantung merupakan gangguan terhadap fungsi kerja jantung. Penyakit ini memiliki banyak jenis seperti kardiovaskuler, jantung koroner dan serangan jantung. Penyakit jantung menjadi salah satu penyakit yang kasusnya sering terjadi di masyarakat tanpa memandang usia, jenis kelamin dan gaya hidup [1]. Penyakit jantung sendiri dapat disebabkan oleh banyak hal, seperti penyumbatan pada pembuluh darah, peradangan, infeksi, ataupun kelainan lainnya. Adapun penyebab seseorang dapat terkena penyakit jantung melibatkan banyak faktor, misalnya terlalu sering mengonsumsi rokok, riwayat keluarga, pola makan yang tidak teratur, tekanan darah yang tinggi, kadar kolesterol yang tinggi, diabetes, tidak menjaga kebersihan tubuh, usia, dan juga jenis kelamin [2]. Melalui faktor diatas, terdapat beberapa hal yang biasanya dirasakan oleh pengidap penyakit jantung yaitu sesak napas, mual, keringat dingin, jantung berdebar, nyeri dada hingga hilangnya kesadaran atau pingsan [3].

Berdasarkan WHO (World Health Organization) yang merupakan sebuah organisasi kesehatan dunia menyatakan bahwa penyakit jantung adalah penyebab utama kematian di Inggris, Amerika Serikat, Kanada dan Australia. Jumlah orang dewasa yang didiagnosis dengan penyakit jantung terdiri dari 26,6 Juta Jiwa (11,3%) populasi orang dewasa [1]. Penyakit jantung merupakan salah satu penyumbang kematian terbanyak di dunia setiap tahunnya, angkanya mencapai sekitar 17 juta orang yang meninggal akibat penyakit jantung [4]. Dengan demikian, diagnosa dini terhadap penyakit jantung sangat penting untuk dilakukan. Namun, prediksi secara manual dapat dikatakan kurang efektif akibat kurangnya keahlian staff medis yang dapat menghasilkan prediksi yang salah [5].

Seiring dengan berkembangnya waktu, teknologi yang semakin maju juga sangat membantu dalam bidang kesehatan terlebih lagi untuk menangani berbagai penyakit. Saat ini telah berkembangnya sebuah teknologi yang disebut juga sebagai kecerdasan buatan. Teknologi ini merupakan sebuah pengembangan sistem atau mesin dalam melakukan hal-hal yang biasanya dilakukan oleh manusia. Maka dari itu, melalui penelitian ini penulis akan memanfaatkan sebuah teknologi kecerdasan buatan yaitu *machine learning*.

Machine learning, merupakan sebuah pendekatan pengolahan data yang terus belajar dan berkembang melalui pengalaman dan data yang diterima. Pada sistem prediksi risiko penyakit jantung yang disusun oleh penulis, akan digunakan metode SVM (*Support Vector Machine*) yang menjadi bagian dari *machine learning*. *Support Vector Machine* sendiri merupakan salah satu algoritma *machine learning* yang termasuk ke dalam kategori *supervised learning*. Metode ini biasanya digunakan untuk kasus klasifikasi dan regresi [6].

Penulis memutuskan untuk menggunakan metode SVM yang diduga dapat dengan baik memproses data yang kompleks dengan fitur yang banyak. Data yang digunakan oleh penulis merupakan dataset dengan judul "HEART DISEASE" yang memiliki 14 fitur dan mencakup 4 wilayah yaitu Cleveland, Hungaria, Switzerland, dan Long Beach V. Dimana keempat wilayah tersebut merupakan wilayah yang berada di Eropa dan Amerika Serikat.

Meneliti lebih lanjut, terdapat sebuah studi komprehensif di setiap kabupaten di Eropa dan Amerika Serikat. Melalui studi tersebut, ditemukannya bahwa 58 persen dari 1.391 daerah di Eropa memiliki lebih banyak kematian dibandingkan kelahiran pada dekade pertama abad ke-21 yakni dalam rentang waktu 2000 hingga 2009). Hasil tersebut lebih banyak dibandingkan dengan hasil di Amerika Serikat yang angka kematiannya hanya 28 persen dari 3.141 kabupaten [7].

Hal yang cukup mengejutkan mendengar banyaknya dan tingginya angka kematian dari negara Eropa dan juga Amerika Serikat, saat ternyata berdasarkan *Bloomberg Healthiest Country Index* edisi 2019, banyaknya wilayah bagian Eropa dan Amerika Serikat menjadi wilayah dengan tingkat kesehatan yang tinggi bahkan Swiss yang berada di wilayah Switzerland pun masuk kedalam urutan kelima [8]. Dengan pola hidup yang ternyata produktif dan juga teratur, ternyata tidak menutup kemungkinan untuk sebuah negara memiliki angka

kematian yang besar akibat penyakit jantung.

Berdasarkan penjabaran latar belakang diatas, penulis ingin membuat sebuah sistem prediksi risiko penyakit jantung untuk keempat wilayah di Eropa dan Amerika Serikat tersebut untuk membantu menurunkan angka kematian dengan cara pencegahan secara dini atas penyakit jantung.

II. LANDASAN TEORI DAN METODE

LANDASAN TEORI

A. Penyakit Jantung

Jantung merupakan salah satu organ penting yang terdapat didalam tubuh dan memiliki tanggung jawab untuk memompa darah dan menyuplai oksigen ke seluruh tubuh [9]. Apabila kinerja jantung terganggu, jantung tidak dapat melakukan tugasnya dengan baik, sehingga pompa darah dan oksigen tidak optimal. Hal ini dapat terjadi karena sebagian otot jantung mati yang disebabkan oleh penyempitan arteri [10] atau terdapat penumpukan plak di arteri koroner [11]. Menurut World Health Organization atau WHO pada tahun 2021 [12], penyakit jantung terdapat beberapa jenis, diantaranya adalah sebagai berikut [13].

- 1) Penyakit jantung koroner merupakan salah satu penyakit jantung yang diakibatkan oleh menumpuknya plak pada pembuluh darah dan menyebabkan jantung tidak dapat memompa darah dengan baik.
- 2) Penyakit serebrovaskular merupakan salah satu penyakit jantung yang diakibatkan oleh adanya penyumbatan yang mengakibatkan pasokan ke otak menjadi terganggu.
- 3) Penyakit arteri perifer merupakan salah satu penyakit jantung yang diakibatkan oleh penyempitan pembuluh darah yang menyebabkan penyumbatan sehingga pasokan darah ke lengan dan kaki menjadi terganggu.
- 4) Penyakit jantung rematik merupakan salah satu penyakit jantung yang disebabkan oleh bakteri streptokokus sehingga menyebabkan terjadinya kerusakan pada otot dan katup jantung.
- 5) Penyakit jantung bawaan merupakan salah satu penyakit jantung yang disebabkan oleh kelainan struktur jantung sejak lahir.
- 6) Trombosis vena dalam dan emboli paru merupakan salah satu penyakit jantung yang disebabkan oleh gumpalan darah di pembuluh darah dan dapat berpindah ke jantung dan paru-paru.

Pada umumnya, penyakit jantung dapat disebabkan oleh dua faktor, yaitu faktor yang tidak dapat diubah dan faktor yang dapat diubah. Faktor yang tidak dapat diubah adalah faktor yang sudah terjadi dan tidak dapat diubah dengan cara apapun. Faktor yang tidak dapat diubah meliputi umur, jenis kelamin, dan faktor genetik. Sedangkan faktor yang dapat diubah merupakan faktor yang sudah terjadi tetapi masih dapat diperbaiki. Faktor yang dapat diubah meliputi gaya hidup, berat badan, hingga kurangnya aktivitas fisik [10].

B. Machine Learning

Machine learning atau pembelajaran mesin merupakan jenis pengaplikasian dari kecerdasan buatan atau *artificial intelligence* yang memiliki kemampuan untuk belajar dari penginputan data [14]. Dalam hal ini, data historis dapat digunakan pada *machine learning* sebagai input untuk melakukan prediksi masa depan [15]. Pada umumnya, *machine learning* dapat digunakan untuk berbagai bidang dengan masalah yang berbeda-beda seperti prediksi risiko hingga pengenalan pola [16]. *Machine learning* memiliki empat kategori pembelajaran, diantaranya sebagai berikut [17].

- 1) *Supervised learning* atau pembelajaran terbimbing merupakan salah satu jenis pembelajaran yang terdapat pada *machine learning* dan digunakan untuk melakukan penyelesaian masalah klasifikasi serta regresi.
- 2) *Semi-supervised learning* atau pembelajaran semi terbimbing merupakan salah satu kategori pembelajaran yang terdapat pada *machine learning* dan sifatnya mirip dengan *supervised learning*, namun yang membedakan antara *supervised learning* dengan *semi-supervised learning* ada pada proses pelabelan data.
- 3) *Unsupervised learning* atau pembelajaran tidak terbimbing merupakan salah satu jenis pembelajaran yang terdapat pada *machine learning* dan digunakan untuk melatih sistem. Untuk melatih sistem, *unsupervised learning* menggunakan data yang tidak memiliki label.
- 4) *Reinforcement learning* merupakan salah satu kategori pembelajaran yang terdapat pada *machine learning* yang digunakan untuk melatih agen untuk belajar melalui *trial and error* dengan menggunakan sistem *reward and punishment* [18].

C. Support Vector Machine

Algoritma *support vector machine* merupakan salah satu jenis algoritma dari *supervised learning* yang digunakan untuk melakukan perbandingan standar nilai diskrit pada parameter serta mengambil salah satu nilai pada parameter yang memiliki akurasi klasifikasi terbaik [19]. Persamaan untuk *decision function* dari algoritma *support vector machine* adalah sebagai berikut [20].

$$w \cdot x + b = 0 \quad (1)$$

Keterangan :

w = parameter *hyperplane* yang dicari
 x = titik data *input support vector machine*
 b = nilai bias

Pada umumnya, algoritma *support vector machine* digunakan untuk menyelesaikan masalah klasifikasi *linear*. Namun seiring berkembangnya waktu, algoritma *support vector machine* dapat melakukan klasifikasi *non-linear* [21]. Selain menyelesaikan masalah klasifikasi *linear* dan *non-linear*, algoritma *support vector machine* juga dapat digunakan untuk menyelesaikan masalah regresi.

1) **Kernel:** Dalam menyelesaikan masalah klasifikasi dan regresi, algoritma *support vector machine* dapat menggunakan *kernel* yang merupakan pemisah antara kelas satu dengan kelas yang lain. Pada umumnya, algoritma *support vector machine* memiliki tiga jenis *kernel*, yang diantaranya sebagai berikut [22].

- 1) *Linear* merupakan salah satu *kernel* yang dapat digunakan dalam algoritma *support vector machine* dengan cara menggunakan garis lurus sebagai *hyperplane* untuk antar kelas. Pada umumnya, *kernel linear* membutuhkan dua jenis variabel yaitu x_i dan x_j . Untuk melakukan perhitungan *support vector machine* dengan menggunakan *kernel linear*, nilai dari x_i akan dilakukan *transpose* dan kemudian dikalikan dengan x_j .

$$k(x_i, x_j) = x_i^T x_j \quad (2)$$

Keterangan:

x_i = input vektor i
 x_j = input vektor j
 x_i^T = transpose dari vektor x_i

- 2) RBF atau *Radial Basis Function* merupakan salah satu *kernel* yang dapat digunakan dalam algoritma *support vector machine* dengan cara menggunakan parameter gamma dan C. Dalam hal ini, gamma memiliki fungsi untuk menjadi batas keputusan dan wilayah keputusan. Nilai gamma yang digunakan harus lebih dari angka 0 sehingga pada umumnya, nilai yang digunakan berkisar dari 0.0001 hingga 10. Sedangkan C memiliki fungsi untuk menjadi penalti terhadap kesalahan ketika melakukan klasifikasi.

$$\exp(-\gamma \|x_i - x\|^2), \gamma > 0 \quad (3)$$

keterangan:

$-\gamma$ = konstanta yang mempengaruhi sensitivitas kernel
 $\|x_i - x\|^2$ = jarak euclidean antara x_i dan x

- 3) *Polynomial* merupakan salah satu *kernel* yang dapat digunakan dalam algoritma *support vector machine* dengan cara menggunakan dua parameter yang berbeda dari *kernel linear* dan *kernel RBF*. *Kernel polynomial* menggunakan parameter r yang merupakan parameter bebas dan parameter d yang merupakan kuadrat.

$$k(x_i, x) = (y \cdot x_i^T x + r)^d \quad (4)$$

Keterangan:

x_i dan x = input vektor
 y = label kelas dari data input
 r = parameter bebas
 d = derajat polinomial

METODE PENELITIAN

D. Tahapan Penelitian

Pada tahapan penelitian ini terdapat gambaran umum alur penelitian yang dilakukan dari awal hingga akhir. Tahapan

penelitian yang dipakai dalam penelitian dijabarkan dalam bentuk *flowchart*. Seluruh tahapan penelitian dilakukan menggunakan bahasa pemrograman python. Berikut adalah gambar dari *flowchart* tahapan penelitian. [23]

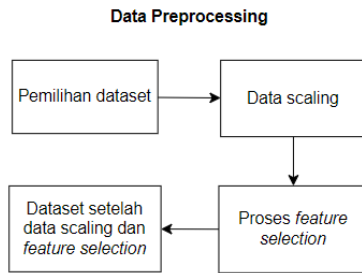


Fig. 1. Flowchart Preprocessing

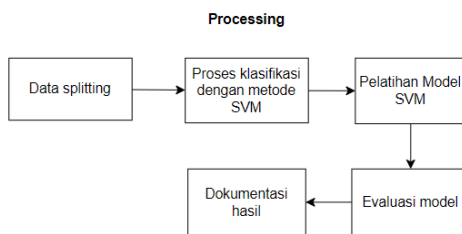


Fig. 2. Flowchart Processing

E. Dataset

Dataset yang digunakan bersumber dari Kaggle <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset> yang didalamnya terdapat 14 fitur yang mencakup usia, jenis kelamin, *chest pain* dan data pendukung prediksi lainnya.

Atribut	Keterangan
Age	Usia
Sex	Jenis Kelamin
Chest Pain Type	Jenis nyeri dada
Cholesterol	Kadar Kolesterol
FastingBS	Gula Darah Puasa
RestingECG	Hasil <i>Electrocardiographic</i> Istirahat
MaxHR	Tingkat Detak Jantung Maksimum
ExerciseAngina	Induksi Angina
Oldpeak	Tingkat Depresi
Slope	kemiringan segmen ST
ca (Vessels)	Jumlah pembuluh darah utama
Thal	Hasil penggambaran talium
Target	Klasifikasi data

TABLE I
ATRIBUT DATASET

F. Preprocessing

Tahap awal dalam penelitian adalah *preprocessing*. Tahap ini dilakukan untuk memodifikasi data yang akan digunakan untuk pelatihan model *machine learning*.

1) **Visualisasi Data:** Visualisasi data adalah proses mewakili data secara grafis, baik dalam bentuk grafik, diagram, atau plot, dengan tujuan untuk membuat data lebih dapat dimengerti, mengidentifikasi pola, dan mengekspresikan informasi yang tersembunyi.

2) **Data Scaling:** *Data scaling* adalah proses mengubah rentang nilai (skala) dari suatu variabel dalam dataset sehingga variabel-variabel tersebut memiliki skala yang seragam. Tujuan utama dari *data scaling* adalah untuk memastikan bahwa variabel-variabel tersebut berada pada skala yang setara, sehingga perbandingan atau perhitungan jarak antara variabel-variabel tersebut dapat dilakukan dengan benar.

3) **Feature Selection:** *Feature selection* adalah proses pemilihan subset fitur dari suatu dataset yang paling relevan atau signifikan untuk digunakan dalam analisis atau model prediksi. Tujuan utama dari *feature selection* adalah untuk meningkatkan kinerja model dengan mengurangi dimensi dataset, sehingga hanya fitur-fitur yang paling informatif atau penting yang digunakan. Di dalam alur penelitian penulis, penulis melakukan seleksi fitur dengan menggunakan *Heatmap Correlation*.

4) **Outlier & Null Value Checking:** Pengecekan *outlier* adalah langkah untuk mengidentifikasi nilai-nilai yang signifikan atau ekstrem dalam dataset. *Outlier* adalah data yang secara signifikan berbeda dari mayoritas data. Sedangkan, Pengecekan nilai *null* atau *missing values* adalah proses untuk mengidentifikasi apakah dataset mengandung nilai yang hilang atau *null*.

G. Processing

Setelah dilakukan tahap data *preprocessing*, selanjutnya akan dimulai tahap *processing* dimana disini dilakukan beberapa hal inti dalam penelitian.

1) **Data Splitting:** *Data splitting* adalah proses membagi dataset menjadi beberapa subset yang berbeda untuk digunakan pada berbagai tahap dalam pembangunan dan evaluasi model *machine learning*. Pemisahan data ini penting untuk mengukur kinerja model secara objektif dan memastikan bahwa model dapat melakukan generalisasi dengan baik pada data yang belum pernah dilihat sebelumnya.

2) **Proses Klasifikasi dengan Metode SVM:** Dalam proses ini, data yang sudah dilakukan *splitting* dan sudah melewati berbagai proses *preprocessing* digunakan untuk proses klasifikasi deteksi penyakit jantung menggunakan algoritma SVM. Klasifikasi ini bertujuan agar *output* yang dihasilkan oleh mesin nantinya akan sesuai dengan kriteria yang ingin ditampilkan. Pada penelitian ini, penulis mengklasifikasikan data menjadi 5 tipe *output*, yang terdiri dari risiko rendah, risiko sedang, risiko tinggi, risiko sangat tinggi dan risiko berat.

3) **Model SVM:** Model SVM yang penulis kembangkan memiliki fitur yang dapat memberi *output* sesuai input dari pengguna, contohnya jika pengguna ingin mengecek data index ke-10 dalam dataset, maka model SVM akan mencari dan mengklasifikasikan apakah pasien dengan index ke-10 tersebut memiliki risiko penyakit jantung di tingkat apa.

4) **Underfitting / Overfitting Checking:** *Underfitting* merupakan keadaan dimana model tidak memiliki kemampuan yang maksimal pada saat melakukan proses pada data pelatihan atau data *train*. Jika ini terjadi, maka model akan menghasilkan kinerja buruk pada data *train*. Sedangkan, *Overfitting* terjadi karena model memiliki kompleksitas yang lebih. Misalnya, model terlalu menghafal pola data sebelumnya dan karena itu, model tidak "mempelajari" pola baru. [24]

H. Tuning Model

Tuning model merupakan proses menyesuaikan parameter atau konfigurasi model untuk meningkatkan kinerja pada set data tertentu. Prosesnya melibatkan pemilihan parameter yang optimal, seperti tingkat pembelajaran, jumlah lapisan dan unit dalam jaringan saraf, atau parameter lainnya, untuk meningkatkan kemampuan model untuk mempelajari pola yang mewakili data dengan baik.

I. Evaluasi

Setelah melakukan *training model SVM*, penulis melakukan evaluasi model yang mencakup akurasi model, nilai *Precision*, nilai *Recall* dan *F1-Score*.

J. Dokumentasi Hasil

Setelah semua proses *machine learning* selesai, penulis melakukan dokumentasi hasil yang akan dijabarkan dalam bab selanjutnya.

III. HASIL DAN DISKUSI

Berikut dilampirkan hasil dari penelitian yang telah dilakukan oleh penulis. Hasil berikut merupakan persentase risiko orang terkait mengenai terkena penyakit jantung dan tidak terkena serta klasifikasi dari risiko tersebut. Probabilitas 0 adalah probabilitas seseorang tidak terkena penyakit jantung dan probabilitas 1 adalah probabilitas seseorang terkena penyakit jantung. Berikut penulis sertakan beberapa hasil output dari setiap index serta besar persentasenya dan klasifikasinya.

[illegible]

Fig. 3. Hasil Perhitungan dalam Tabel

Setelah dilakukan pencetakan hasil output untuk memprediksi dari setiap index dalam bentuk tabel, berikutnya ditampilkan dalam format lain. Pada tahap ini, user dapat melakukan input sesuai yang ingin dicari dengan melakukan input nomor index dan program akan menentukan untuk persentase serta klasifikasi dari risiko index yang diinput oleh user. Berikut penulis sertakan potongan kode dan hasil keluarannya.

[illegible]

Fig. 4. Hasil Perhitungan menggunakan Input User

Setelah mencetak hasil dari *input user*, selanjutnya yaitu menghitung nilai akurasi, presisi, *f1 score*, dan *recall* dari model SVM yang digunakan. Didapatkan hasil akurasi dari model SVM pada prediksi penyakit jantung adalah sebesar 91 persen. Berikut penulis sertakan potongan kode dan hasil dari akurasi, presisi, *f1 score*, dan *recall* model SVM pada prediksi penyakit jantung.

```

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.05, random_state=42)

model = StandardScaler()
x_train_scaled = model.fit_transform(x_train)
x_test_scaled = model.transform(x_test)

sum_model = LogisticRegression()
sum_model.fit(x_train_scaled, y_train)

predict_sum = sum_model.predict(x_test_scaled)

print('sum = accuracy: %.2f, f1: %.2f' % (accuracy_score(y_test, predict_sum), f1_score(y_test, predict_sum)))
print('sum = precision: %.2f, recall: %.2f' % (precision_score(y_test, predict_sum), recall_score(y_test, predict_sum)))
print('sum = f1_score: %.2f' % (f1_score(y_test, predict_sum)))

print('sum = accuracy: %.2f, f1: %.2f' % (accuracy_score(y_test, predict_sum), f1_score(y_test, predict_sum)))
print('sum = precision: %.2f, recall: %.2f' % (precision_score(y_test, predict_sum), recall_score(y_test, predict_sum)))
print('sum = f1_score: %.2f' % (f1_score(y_test, predict_sum)))

metrics_model_sum = f1_score(y_test, predict_sum)

Recall_sum = 0.000
Precision_sum = 0.000
F1_sum = 0.000

```

Fig. 5. Akurasi, Precision, Recall, dan F1 Score Model SVM

Pengukuran kinerja dalam model SVM yang digunakan dalam memprediksi jantung menggunakan *confusion matrix*. Hasil menunjukkan bahwa pada dataset yang bernilai *false* dan diprediksi *false* sebanyak 42. yang bernilai *false* namun diprediksi *true* sebesar 14. Kemudian, yang bernilai *true* dan diprediksi *true* sebesar 44 dan yang bernilai *true* serta diprediksi *false* sebesar 1.

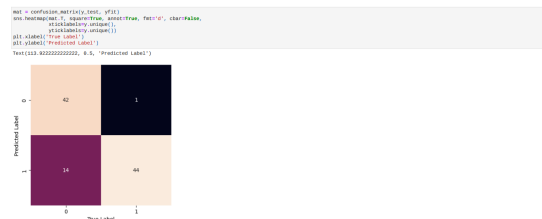


Fig. 6. *Confusion Matrix*

Model *Support Vector Machine* yang penulis kembangkan memiliki output yang menentukan pasien tersebut memiliki

probabilitas 0 (tidak terkena penyakit jantung) atau prob. 1 (terkena penyakit jantung), output didapat dari hasil kalkulasi sesuai index dataset yang dipilih. Contoh pasien index ke-1 dalam dataset memiliki Probabilitas 1 lebih tinggi sebesar 0.9169 dan Probabilitas 0 sebesar 0.0831, maka pasien index ke-1 ini memiliki risiko terkena penyakit jantung. Dalam output penelitian juga ditampilkan tingkat risiko berdasarkan nilai output pada probabilitas 1 dan 0 dengan contoh pasien index ke-1 termasuk dalam kategori “risiko Berat”.

Pada penelitian prediksi penyakit jantung menggunakan algoritma *Support Vector Machine* ini dihasilkan tingkat akurasi model *Support Vector Machine* yang digunakan sebesar 91.09 persen, presisi sebesar 86 persen, *recall* sebesar 95.56 persen, dan *F1 score* sebesar 90.53 persen. Dari hasil perhitungan metrik evaluasi tersebut, penulis dapat mengatakan bahwa penelitian ini cocok untuk melakukan prediksi penyakit jantung dengan algoritma dan dataset yang telah digunakan.

penulis melakukan perbandingan dengan algoritma lain yang penelitiannya dilakukan oleh Victor Chang, Vallabhant Rupa Bhavani, Qianwen Xu, dan Alamgir Hosain yang berjudul “An artificial intelligence model for heart disease detection using machine learning algorithms”, dihasilkan akurasi dari model algoritma yang berbeda, dataset dan ukuran tes yang sama yaitu sebanyak 100 data. Penelitian yang mereka lakukan memiliki hasil akurasi sebesar 87 persen untuk algoritma kNN, 79 persen untuk algoritma *decision tree*, 84 persen untuk algoritma *random forest*, dan sebesar 83 persen untuk algoritma SVM [25].

Perbandingan dengan penelitian yang dilakukan oleh Perbandingan dengan penelitian yang dilakukan oleh Dwi Sidik Permana dan Astried Silvanie dalam jurnal yang berjudul “Prediksi Penyakit Jantung menggunakan Support Vector Machine dan Python pada Basis Data Pasien di Cleveland” dihasilkan bahwa akurasi penelitian menunjukkan angka sebesar 90.11 persen [26].

Dari beberapa hasil penelitian, menunjukkan keberagaman hasil akurasi dari berbagai algoritma yang digunakan dalam penelitian. Kebanyakan dari hasil akurasi penelitian penelitian tersebut menunjukkan bahwa model algoritma SVM cocok karena menunjukkan bahwa angka akurasi tersebut selalu tinggi.

IV. KESIMPULAN

Melalui penelitian penulis, penggunaan algoritma SVM menggunakan data set “Heart Disease Dataset” dari situs kaggle, didapatkan dengan adanya 14 fitur yang berada dalam dataset. Fitur-fitur tersebut ialah *age*, *sex*, *cp*, *trestbps*, *chol*, *fbs*, *restecg*, *thalach*, *exang*, *oldpeak*, *slope*, *ca*, *thal*, dan *target*. 14 fitur inilah yang dapat mempengaruhi akurasi serta penentuan prediksi risiko penyakit jantung yang dimiliki. Pada penelitian ini penulis membuat 5 tingkatan skala risiko, yaitu, ‘Risiko Rendah’, ‘Risiko Sedang’, ‘Risiko Tinggi’, ‘Risiko Sangat Tinggi’, dan ‘Risiko Berat’.

Setelah melakukan perhitungan menggunakan metode SVM, dari 100 orang yang di tes melalui riwayat kesehatannya, terdapat 51 yang memiliki ‘Risiko Berat’, 2 yang memi-

liki ‘Risiko Sangat Tinggi’, 2 yang memiliki ‘Risiko Tinggi’, 3 yang memiliki ‘Risiko Sedang’, dan 43 yang memiliki ‘Risiko Rendah’.

Setelah melakukan penentuan risiko, akhirnya penulis mengukur akurasi dari algoritma SVM ini menggunakan dataset yang sama. Dimana hasil akurasinya mencapai hingga angka 91

Melihat dari hasil keseluruhan, dengan adanya 51 orang yang memiliki risiko berat, 51 orang tersebut dapat mencegah secara dini, agar risiko terkena penyakit jantung semakin menurun. Sehingga, tujuan penelitian penulis dapat dinyatakan berhasil.

REFERENCES

- [1] D. P. Utomo and M. Mesran, “Analisis komparasi metode klasifikasi data mining dan reduksi atribut pada data set penyakit jantung,” *Jurnal Media Informatika Budidarma*, 2020.
- [2] dr. Pittara, “Penyakit jantung,” 16 Maret 2023. [Online]. Available: <https://www.alodokter.com/penyakit-jantung/penyebab>
- [3] dr. Sienny Agustin, “Ciri-ciri sakit jantung yang harus diwaspadai,” 8 Mei 2022. [Online]. Available: <https://www.alodokter.com/Kenali-Ciri-Ciri-Sakit-Jantung>
- [4] W. H. Organization, “Who director-general’s opening remarks at the second annual gathering for the global ncd compact,” 21 September 2023. [Online]. Available: <https://www.who.int/director-general/speeches/detail/>
- [5] V. S. S. G. D. P. G. A. H. . G. J. Pouriyeh, S., “A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease,” 2017.
- [6] I. R. I. A. Ary Putranto1, Nuril Lutvi Azizah, “Sistem prediksi penyakit jantung berbasis web menggunakan metode svm dan framework streamlit,” *Jurnal Penerapan Sistem Informasi (Komputer Manajemen)*, p. 11, 2023.
- [7] Q. R. . W. D. Putri, “Angka kematian di eropa lebih banyak dibanding angka kelahiran,” 19 Jan 2016.
- [8] L. J. Miller and W. Lu, “These are the world’s healthiest nations,” *Bloomberg Journal*, 2019.
- [9] B. P. SANTOSO, “Heart rate monitor,” 10 2018.
- [10] D. G. Pradana, M. L. Alghifari, M. F. Juna, and S. D. Palaguna, “Klasifikasi penyakit jantung menggunakan metode artificial neural network,” *Indonesian Journal of Data and Science (IJODAS)*, vol. 3, pp. 55–60, 2022.
- [11] L. Ghani, M. D. Susilawati, and H. Novriani, “Faktor risiko dominan penyakit jantung koroner di indonesia,” *Buletin Penelitian Kesehatan*, vol. 44, 12 2016.
- [12] “Cardiovascular diseases (cvds).” [Online]. Available: [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- [13] N. W. J. NARYADI, “Hubungan tingkat pengetahuan, tingkat dukungan keluarga dan tingkat kepatuhan diet pasien jantung pasca rawat inap di rumah sakit umum bangli,” 2019.
- [14] A. S. RAHARJO, “Perbandingan klasifikasi serangan jaringan distributed denial of service (ddos) menggunakan algoritma decision tree, k-nearest neighbors, bayesian network dan oner,” 2020.
- [15] M. James and R. Dennis, “Machine-learning and statistical methods for ddos attack detection and defense system in software defined networks,” 5 2021. [Online]. Available: <https://www.semanticscholar.org/paper/Machine-learning-and-statistical-methods-for-DDoS-Dennis/033a589571860de1b499ceab4a81a1c8da646de4>
- [16] A. G. Budianto and A. Syarief, “Analisis pengaruh pengurangan dimensi data pada keakuratan prediksi penyakit jantung dengan menggunakan svm linear,” *Jurnal Taguchi : Jurnal Ilmiah Teknik dan Manajemen Industri*, vol. 3, pp. 80–91, 7 2023. [Online]. Available: <https://taguchi.lppmbinabangsa.id/index.php/home/article/view/58>
- [17] A. Hermawan, “Konsep dasar machine learning dan neural network dalam studi literatur — pdf.” [Online]. Available: <https://id.scribd.com/document/627739208/44>

- [18] J. Andreanus, A. Kurniawan, K. M. V. D. Maitreya, S. Panas, and K. R. Indonesia, "Sejarah, teori dasar dan penerapan reinforcement learning: Sebuah tinjauan pustaka," *Jurnal Telematika*, vol. 12.
- [19] F. F. Handayanna, "Penerapan metode support vector machine menggunakan optimasi genetic algorithm untuk prediksi penyakit diabetes," *Jurnal Teknik Informatika*, vol. 1, pp. 139–147, 2015. [Online]. Available: <https://www.neliti.com/publications/495039/>
- [20] S. Y. Pangestu, Y. Astuti, and L. D. Farida, "Algoritma support vector machine untuk klasifikasi sikap politik terhadap partai politik indonesia," *Jurnal Mantik Penusa*, vol. 3, pp. 236–241, 2019. [Online]. Available: <https://t.co/eF>
- [21] E. Tasia, R. Zaid, I. Z. Ismail, S. Kenia, P. Loka, Y. Ikhsani, and R. Ocviani, "Sentimas: Seminar nasional penelitian dan pengabdian masyarakat classification of heart failure disease using supervised learning klasifikasi penyakit gagal jantung menggunakan supervised learning." [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [22] A. S. Nugraha and K. K. Purnamasari, "Penerapan metode support vector machine pada part of speech tag bahasa indonesia."
- [23] A. Putranto, N. L. Azizah, I. Ratna, I. Astutik, F. Sains, and D. Teknologi, "Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit," *Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 4, no. 2, pp. 442–452, 2023. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/heart+disease>
- [24] W. A. Firmansyach, U. Hayati, and Y. Arie Wijaya, "Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 262–269, 2023.
- [25] V. Chang, V. Bhavani, Q. Xu, and A. Hossain, "An artificial intelligence model for heart disease detection using machine learning algorithms," *Healthcare Analytics*, vol. 2, p. 100016, 01 2022.
- [26] D. S. Permana and A. Silvanie, "Prediksi penyakit jantung menggunakan support vector machine dan python pada basis data pasien di cleveland," vol. 2, pp. 29–34, 4 2021.