

Web-based Heart Disease Prediction System Using SVM Method and Streamlit Framework

[Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit]

Ary Putranto^{1,*}, Nuril Lutvi Azizah², Ika Ratna Indra Astutik³

¹⁾Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

³⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: 201080200191@umsida.ac.id

Abstract. *Heart disease is a disease whose cases often occur in society regardless of age, gender and lifestyle. Most of these diseases can not be cured completely. Heart attacks that are handled too late can cause dangerous complications with the most fatal risk, namely death. With the rapid development of technology at this time, the world of medicine is very helpful. One of the technological advances that exist today is a system that can diagnose disease. One of the machine learning algorithms that can be used is the Support Vector Machine. SVM can be used for regression and classification cases. The advantage of this algorithm is its fast computation. Classification of heart disease with the SVM algorithm results in an accuracy of 85%. The machine learning model that produces the accuracy value is then carried out by deploying the model using the Streamlit framework. Streamlit is a Python-based framework that is open source. This framework was created to make it easier for developers to build web-based programs in the fields of interactive data science and machine learning. The results of this study are web-based programs that can diagnose heart disease with an accuracy value of 85%.*

Keywords - Classification; Heart Disease; Streamlit; SVM

Abstrak. Penyakit jantung merupakan penyakit yang kasusnya sering terjadi di masyarakat tanpa memandang usia, jenis kelamin dan gaya hidup. Penyakit ini Sebagian besar tidak dapat disembuhkan secara total. Serangan jantung yang terlambat dilakukan penanganan dapat menyebabkan komplikasi berbahaya dengan resiko yang paling fatal yaitu kematian. dengan adanya perkembangan teknologi yang sangat pesat saat ini, dunia kedokteran menjadi sangat terbantu. Salah satu kemajuan teknologi yang ada saat ini adalah sistem yang dapat mendiagnosa penyakit. Salah satu algoritma machine learning yang dapat digunakan adalah Support Vector Machine. SVM dapat digunakan untuk kasus regresi dan klasifikasi. Kelebihan dari algoritma ini adalah komputasinya yang cepat. Klasifikasi penyakit jantung dengan algoritma SVM mendapatkan hasil akurasi sebesar 85%. Model machine learning yang menghasilkan nilai akurasi tersebut selanjutnya dilakukan deployment model menggunakan framework Streamlit. Streamlit merupakan framework berbasis Python yang besifat open source. Framework ini dibuat untuk memudahkan developer dalam membangun program berbasis web di bidang data science dan machine learning yang interaktif. Hasil dari penelitian ini adalah program berbasis web yang dapat mendiagnosa penyakit jantung dengan nilai akurasi sebesar 85%.

Kata Kunci - Klasifikasi; Penyakit Jantung; Streamlit; SVM

I. PENDAHULUAN

Penyakit jantung merupakan gangguan terhadap fungsi kerja jantung. Penyakit ini memiliki banyak jenis seperti kardivaskuler, jantung coroner dan serangan jantung. Penyakit jantung menjadi salah satu penyakit yang kasusnya sering terjadi di masyarakat tanpa memandang usia, jenis kelamin dan gaya hidup [1]. Adapun penyebab seseorang terkena serangan jantung yaitu karena kurangnya aktivitas, merokok, stress berat, tingginya kolesterol, obesitas dan lainnya. Beberapa hal yang dialami oleh seseorang yang terkena serangan jantung ialah nyeri dada yang merambat hingga bahu, leher, rahang dan punggung. Pernafasan juga terasa berat dengan diikuti keringat dingin, rasa mual dan hilang kesadaran [2]. Sebagian besar penyakit jantung tidak dapat disembuhkan secara total. Serangan jantung yang parah atau terlambat dilakukan penanganan dapat menyebabkan komplikasi berbahaya dengan resiko yang paling fatal adalah kematian.

Pada tahun 2012, data kasus penyakit di Indonesia berdasarkan GLOBOCAN (IARC) terdapat jumlah kasus baru serangan jantung sebanyak 43,30% dan jumlah kematian sebesar 12,90%. Sedangkan pada tahun 2013 jumlah penderita penyakit jantung sebesar 61.682 orang, dengan penderita terbanyak berada di Jawa Tengah dengan jumlah kasus sebesar 11.511 kasus. Berdasarkan data di atas, dapat diketahui bahwa banyak orang yang belum menanggapi penyebab penyakit ini secara serius. Sehingga setelah melakukan pemeriksaan, dokter mendeteksi adanya penyakit jantung dengan stadium yang tinggi [3].

Perkembangan teknologi yang kian pesat sangat membantu dunia kedokteran dalam menangani berbagai macam penyakit. Saat ini telah ada teknologi yang biasa disebut dengan kecerdasan buatan. Salah satu manfaat dari teknologi ini yaitu dapat memprediksi penyakit dengan tingkat akurasi tertentu. Penelitian ini bertujuan untuk melakukan prediksi dini terhadap indikasi penyakit jantung yang di deployment ke dalam sebuah sistem berbasis web.

Dengan memasukkan beberapa indikator yang menjadi penyebab serangan jantung, program akan memprediksi apakah hasil masukan tersebut terindikasi penyakit jantung atau tidak. Salah satu algoritma machine learning yang dapat diterapkan untuk melakukan klasifikasi adalah Support Vector Machine (SVM).

Support Vector Machine merupakan salah satu algoritma machine learning yang termasuk ke dalam kategori supervised learning. Metode ini biasanya digunakan untuk kasus klasifikasi dan regresi [4]. Pada dasarnya SVM merupakan representasi dari kelas yang berbeda dalam ruang multidimensi. Tujuannya adalah untuk membagi dataset ke dalam kelas-kelas untuk menemukan hyperplane marginal maksimum. Adapun kelebihan dari metode SVM adalah memiliki daya komputasi yang cepat.

Model machine learning yang telah dibuat dapat dimasukkan ke dalam program berbasis web agar pengguna dapat dengan mudah melakukan klasifikasi. Adapun salah satu framework yang mendukung untuk melakukan deployment model ke dalam program berbasis web yaitu streamlit. Streamlit merupakan framework berbasis Python yang besifat open source. Framework ini dibuat untuk memudahkan developer dalam membangun program berbasis web di bidang data science dan machine learning yang interaktif [5]. Salah satu kelebihan dari streamlit adalah developer tidak perlu mengatur tampilan website dengan CSS, HTML dan javascript karena framework streamlit telah menyediakannya melalui fungsi-fungsi yang terdapat pada framework tersebut.

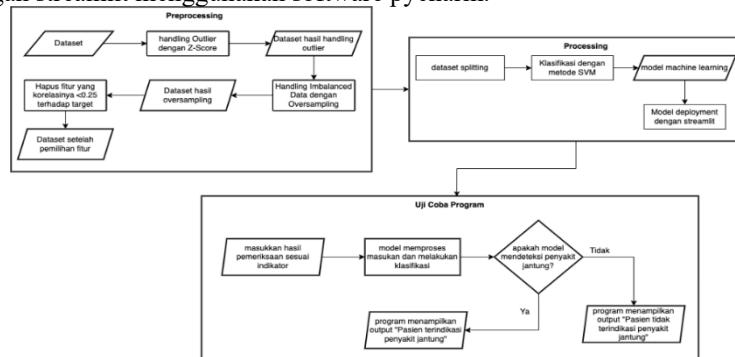
Beberapa penelitian terdahulu telah melakukan penelitian terkait prediksi penyakit jantung. Diantaranya penelitian yang berjudul “Analisa Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Dataset Penyakit Jantung” yang dilakukan oleh Dito Putro Utomo dan Mesran pada tahun 2020. Penelitian ini melakukan reduksi atribut menggunakan metode PCA dengan membandingkan algoritma C5.0 dan Naïve Bayes Classifier untuk klasifikasi. Hasil dari penelitian ini menyatakan bahwa algoritma Naïve Bayes Classifier lebih baik daripada algoritma C5.0 [1].

Berdasarkan latar belakang dan penjelasan yang telah dijabarkan di atas, sehingga penelitian ini akan melakukan prediksi dini terhadap penyakit jantung menggunakan metode Support Vector Machine yang dimasukkan ke dalam program berbasis web sehingga harapan dari penelitian ini yaitu dapat membantu dunia kedokteran untuk melakukan diagnosa penyakit jantung sedini mungkin agar tingkat keparahan dapat diminimalisir.

II. METODE

A. Tahapan Penelitian

Tahapan penelitian merupakan gambaran umum terkait alur penelitian yang akan dilakukan dalam pelaksanaan penelitian ini dari awal hingga akhir. Tahapan yang dilakukan dalam penelitian ini dapat dipaparkan melalui diagram alir seperti pada Gambar 1. Seluruh tahapan pada penelitian ini dikerjakan menggunakan Bahasa pemrograman python. dengan rincian untuk tahapan preprocessing hingga pemodelan menggunakan software jupyter notebook dan untuk deployment dengan streamlit menggunakan software pycharm.



Gambar 1 Gambaran Tahapan Penelitian

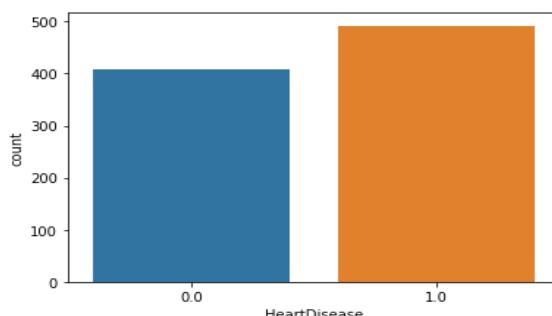
B. Data

Data yang digunakan pada penelitian ini diambil dari <https://archive.ics.uci.edu/ml/datasets/heart+disease> yang terdiri dari 918 rekaman. Dataset ini terdiri dari 12 indikator penyebab penyakit jantung. Berikut merupakan rincian indikator atau attribut dari dataset yang akan digunakan.

Tabel 1 Attribut Dataset

Attribut	Keterangan
Age	Umur
Sex	Jenis kelamin

Chest Pain Type	Jenis nyeri dada
RestingBP	Tekanan Darah
Cholesterol	Kolesterol
FastingBS	Gula Darah Puasa
RestingECG	Hasil Electrodiagram istirahat
MaxHR	Detak jantung maksimum tercapai
ExerciseAngina	Induksi angina
Oldpeak	Depresi
HearDisease	Target (terindikasi penyakit jantung atau tidak terindikasi penyakit jantung)



Gambar 2 Perbandingan Jumlah Rekaman

Gambar 2 merupakan perbandingan jumlah rekaman pasien terindikasi penyakit jantung atau tidak terindikasi penyakit jantung. Nilai 0 merupakan pasien yang tidak terindikasi penyakit jantung dan nilai 1 merupakan pasien yang terindikasi penyakit jantung.

C. Preprocesssing

Tahapan awal penelitian yaitu preprocessing. Preprocessing dilakukan karena data mentah tidak dapat digunakan langsung oleh sistem. Oleh karena itu, beberapa preprocessing harus dilakukan untuk sedikit memodifikasi data guna meningkatkan kualitas data yang digunakan.

Handling Outlier

Sebuah data biasanya terdapat outlier. Outlier merupakan objek pengamatan yang memiliki nilai ekstrim yang terlalu rendah atau terlalu tinggi sehingga menyebabkan masalah dalam penentuan hasil akhir analisis [10]. Terkadang outlier memiliki nilai yang penting dan berpengaruh dalam kumpulan objek, sehingga tidak semua outlier dapat dihilangkan. Oleh karena itu pada penelitian ini, peneliti melakukan handling outlier menggunakan teknik Z-score. Z-score dapat digunakan untuk membantu menentukan apakah sebuah data bernilai ekstrim, atau outlier [11]. Aturan umumnya adalah Z-score dengan nilai kurang dari -3 atau lebih dari +3 menunjukkan bahwa nilai data adalah nilai ekstrim. Sehingga data yang melebihi batas bawah dan batas atas akan dihapus. berikut merupakan Langkah-langkah untuk memfilter outlier dengan Teknik zscore.

1. Menghitung nilai zscore di setiap kolomnya menggunakan library numpy dan fungsi stats pada library scipy. nilai zscore di setiap kolomnya diabsolutkan dan dimasukkan ke dalam variable z.
2. Tahap selanjutnya yaitu memfilter data yang nilai absolutnya kurang dari 3.

Cek Correlation

Hubungan antar variable merupakan suatu hal yang sangat penting dan berpengaruh terhadap suatu penelitian. Oleh karena itu, pada tahap preprocessing ini, peneliti melakukan pengecekan korelasi data menggunakan fungsi dari library pandas yaitu corr yang divisualisasikan oleh library plotly. Untuk melihat interpretasi korelasi antar dua variable, berikut kriteria hasil perhitungan yang dikutip dari website accurate.id

Tabel 2 Koefisien Korelasi

Koefisien	Keterangan
0	Tidak ada korelasi antara dua variable
>0 – 0,25	Korelasi sangat lemah
>0,25 - 0,5	Korelasi cukup

>0,5 – 0,75	Korelasi kuat
1	Korelasi hubungan sempurna positif
-1	Korelasi hubungan sempurna negatif

Berdasarkan tabel 2 di atas, pada penelitian ini koefisien korelasi data yang lemah akan dihapus sehingga menyisakan korelasi data cukup, kuat dan sempurna.

Handling Imbalanced Data

Pada gambar 2 dapat diketahui bahwa terdapat ketidak seimbangan jumlah rekaman data antara pasien yang terindikasi penyakit jantung dan tidak terindikasi penyakit jantung. Adanya ketidak seimbangan jumlah data ini dapat mempengaruhi model machine learning yang dihasilkan. Sehingga walaupun nilai akurasi yang dihasilkan oleh model cukup tinggi, namun model tidak dapat membedakan apakah pasien terindikasi penyakit jantung atau tidak [12]. Untuk menangani hal tersebut, peneliti melakukan Teknik oversampling menggunakan SMOTE. Metode SMOTE diusulkan pertama kali oleh Chawla pada tahun 2002, dimana kelas minoritas dilakukan oversampling dengan cara membuat “data training sintetis”. Pembuatan data training sintetis tersebut dibuat berdasarkan k-nearest neighbor [13].

D. Processing

Setelah data melalui tahapan preprocessing, selanjutnya adalah data dimasukkan ke tahap pemrosesan sehingga mendapatkan hasil yang diharapkan.

Klasifikasi dengan Support Vector Machine

Support Vector Machine merupakan algoritma klasifikasi supervised learning dimana algoritma ini dapat melakukan prediksi kelas berdasarkan pola yang dihasilkan dari proses training [14]. Klasifikasi SVM menciptakan garis pembatas yang digunakan untuk memisahkan antar kelas. Berikut merupakan Langkah-langkah dari klasifikasi dengan SVM

1. Membagi data menjadi dua bagian yaitu data latih dan data uji dengan persentase perbandingan 80% untuk data latih dan 20% untuk data uji.
2. Tuning parameter untuk mendapatkan parameter yang paling optimal menggunakan Teknik GridSearch Cross Validation yang mana GridSearchCV adalah bagian dari modul scikit-learn yang bertujuan secara otomatis dan sistematis melakukan validasi beberapa model dan setiap hyperparameter [15]. Ketika proses running GridSearchCV sudah selesai, maka akan didapatkan model beserta skor uji dan skor latih. Adapun hyperparameter yang digunakan pada penelitian adalah sebagai berikut.

Tabel 3 Hyperparameter Tuning

Parameter	Value	Keterangan
C	0.1, 1, 10, 100, 1000	Parameter yang bekerja untuk menghindari misklasifikasi di setiap sampel
Gamma	1, 0.1, 0.01, 0.001, 0.0001	Digunakan untuk menentukan seberapa jauh pengaruh dari satu sampel data latih
Kernel	Polynomial	Untuk memisahkan dataset yang bentuknya non-linear

E. Evaluasi

Tahapan evaluasi digunakan untuk mengukur performa model yang dihasilkan. Peneliti menggunakan confusion matrix untuk melakukan analisa performa dari model. Confusion matrix merupakan tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual [16]. Selanjutnya dari confusion matrix tersebut dapat dihitung nilai akurasi, precision, recall dan F1 Score. Peneliti juga menggunakan Mean Absolute Error untuk mengetahui prosentase kesalahan rata-rata mutlak [17]. Mean Absolute Error merupakan dua diantara banyak metode yang digunakan untuk mengukur tingkat keakuratan suatu model peramalan [18].

F. Deployment

Setelah melalui tahapan modeling dan evaluasi, selanjutnya yaitu tahapan deployment model. Dimana model disimpan dalam ekstensi pickle untuk dimasukkan ke dalam sistem. Framework yang digunakan pada pembuatan sistem adalah streamlit. berikut merupakan desain sistem berbasis web beserta contoh pengisiannya.

Prediksi Penyakit Jantung

Masukkan Umur

Masukkan Jenis Nyeri Dada

Masukkan Fasting Blood Sugar

Masukkan Detak Jantung Maksumal

Apakah Mengalami Angin Duduk?
Masukkan 1 jika ya, masukkan 0 jika tidak

Masukkan Oldpeak

Masukkan Kemiringan Detak Jantung

Apakah Pasien Perempuan?
Masukkan 1 jika ya, masukkan 0 jika tidak

Apakah Pasien Laki-laki?
Masukkan 1 jika ya, masukkan 0 jika tidak

Prediksi Penyakit Jantung

Gambar 3 Desain Sistem Berbasis Web

G. Pengujian dengan Black Box

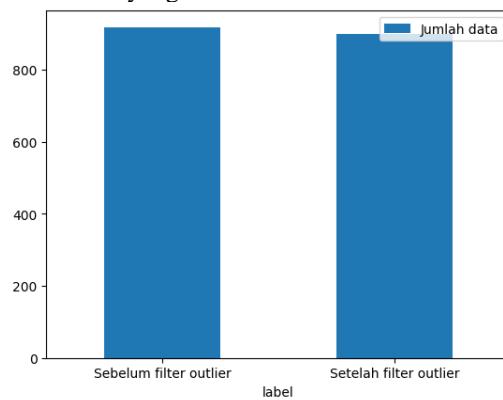
Tahapan selanjutnya yaitu pengujian program berbasis web. Peneliti menggunakan Teknik pengujian black box. Pengujian black box merupakan pengujian yang dilakukan untuk mengamati hasil input dan output dari perangkat lunak tanpa melihat struktur kodennya. Pengujian ini dilakukan di tahapan akhir pada proses pembuatan perangkat lunak dengan tujuan untuk mengetahui apakah perangkat lunak dapat berfungsi dengan baik. Teknik yang digunakan pada black box testing adalah Teknik equivalence partitions. Teknik ini merupakan suatu pengujian yang didasarkan pada masukan data pada setiap form yang telah disediakan [19]. Setiap menu akan pengujiannya dan dikelompokkan berdasarkan fungsinya. Baik itu hasilnya diterima atau tidak diterima.

III. HASIL DAN PEMBAHASAN

A. Hasil Preprocessing

Hasil Handling Outlier

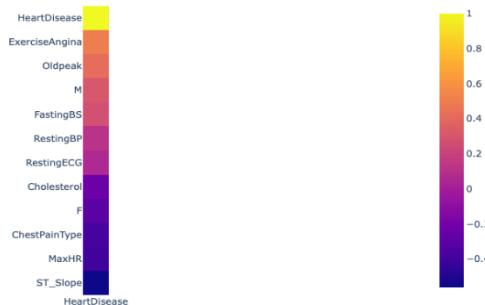
Seperti yang telah dijelaskan pada sub bab 2.3.1 bahwa tahapan awal yang dilakukan yaitu melakukan filter outlier dengan Teknik Z-Score. Setelah kode sumber 2 dijalankan, data yang awalnya 918 record berubah menjadi 899 record. Artinya terdapat 19 record merupakan outlier yang melebihi batas atas dan batas bawah.

*Gambar 4 Hasil Filter Outlier*

B. Hasil Cek Correlation

Setiap indikator memiliki hubungan positif atau negative dengan target, namun tidak semua indikator memiliki hubungan yang kuat terhadap target. sehingga hubungan yang lemah dapat mempengaruhi hasil akurasi. Sehingga pada tahapan selanjutnya yang dilakukan oleh peneliti yaitu tahapan cek correlation. Tahapan ini digunakan untuk

menentukan manakah variable yang berpengaruh dan tidak berpengaruh terhadap target. Pada penelitian ini, peneliti menghapus variable yang memiliki korelasi lemah terhadap target dimana parameter korelasinya adalah di atas -0.25 (korelasi negatif lemah) dan 0.25 (korelasi positif lemah).

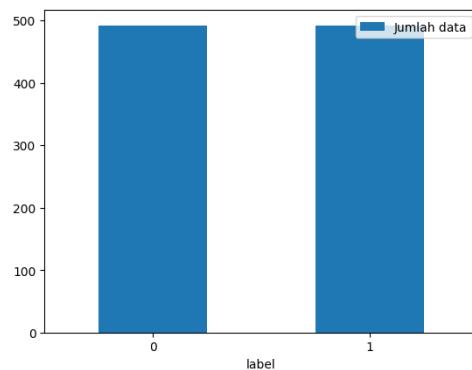


Gambar 5 Visualisasi Hubungan Antar Variabel

Berdasarkan gambar 5 diatas, dapat diketahui bahwa variable Cholesterol, RestingECG dan RestingBP memiliki korelasi yang lemah terhadap target dimana koefisien dari ketiga indikator tersebut di atas -0.25 dan di bawah 0.25. Sehingga peneliti menghapus ketiga variable tersebut.

C. Hasil Handling Imbalanced Data

Tahapan selanjutnya yaitu handling imbalanced data. Pada penelitian ini, peneliti menggunakan SMOTE untuk melakukan oversampling terhadap kelas minoritas. Dapat diketahui pada gambar 2, kelas 0 (tidak terindikasi penyakit jantung) memiliki jumlah kelas yang lebih sedikit dibandingkan kelas 1 (terindikasi penyakit jantung). Sehingga kelas 0 dilakukan oversampling menggunakan SMOTE.



Gambar 6 Hasil Over Sampling

Adapun hasil dari handling imbalanced data yang dilakukan pada kode sumber 3 menghasilkan jumlah kelas yang seimbang yaitu kelas 0 (tidak terindikasi penyakit jantung) adalah 492 record dan kelas 1 (terindikasi penyakit jantung) adalah 492 record.

Processing

Setelah melalui tahapan preprocessing, selanjutnya yaitu tahapan processing. Berdasarkan kode sumber 1, dataset dibagi menjadi 80% untuk data latih dan 20% untuk data uji. Selanjutnya dilakukan hyperparameter tuning menggunakan Grid Search Cross Validation dengan list parameter yang terdapat pada tabel 3. Grid Search Cross Validation merupakan bagian dari scikit-learn yang berfungsi secara otomatis dan sistematis dalam melakukan validasi beberapa model dan setiap hyperparameter. Berdasarkan proses tersebut, dihasilkan parameter terbaik sebagai berikut.

Tabel 4 Hasil Hyperparameter Tuning

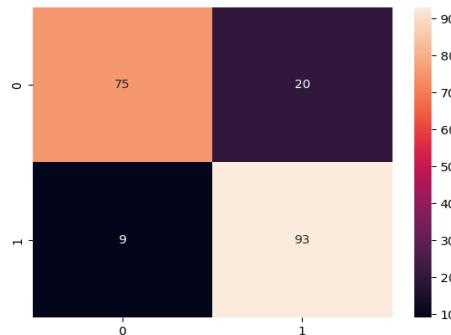
Parameter	Value
C	0.1
Gamma	0.01
Kernel	Poly

Parameter yang terdapat pada tabel 4 selanjutnya dilakukan fitting model dengan memasukkan algoritma Support Vector Machine, berdasarkan uji coba tersebut dihasilkan skor latih sebesar 85% dan skor uji sebesar 87%. Sehingga model yang dihasilkan fit. Kernel terbaik yang dihasilkan oleh Grid Search Cross Validation yaitu poly atau

polynomial. Hal itu dikarenakan persebaran data yang divisualisasikan pada gambar 3 merupakan distribusi data yang tidak linear. Sehingga kernel polynomial memang cocok digunakan untuk penelitian ini.

Evaluasi

Setelah tahapan processing, tahapan selanjutnya yaitu evaluasi terhadap model. Hal ini bertujuan untuk mengukur kinerja dari model yang telah dibuat. pada penelitian ini, peneliti menggunakan confusion matrix. Confusion matrix dapat memberikan informasi perbandingan hasil prediksi dengan hasil klasifikasi yang sebenarnya [20]. Adapun hasil visualisasi confusion matrix pada penelitian ini adalah sebagai berikut.



Gambar 7 Hasil Confusion Matrix

Berdasarkan hasil confusion matrix di atas, dapat dilakukan perhitungan precision, recall, dan accuracy. Berikut merupakan rumus beserta perhitungannya.

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (1)$$

$$Akurasi = \frac{75 + 93}{75 + 9 + 20 + 93} \times 100\%$$

$$Akurasi = \frac{168}{197} \times 100\%$$

$$Akurasi = 0,853 \times 100\%$$

$$Akurasi = 85\%$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Precision = \frac{75}{75 + 20} \times 100\%$$

$$Precision = \frac{75}{95} \times 100\%$$

$$Precision = 0,78 \times 100\%$$

$$Precision = 78\%$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Recall = \frac{75}{75 + 9} \times 100\%$$

$$Precision = \frac{75}{84} \times 100\%$$

$$Precision = 0,89 \times 100\%$$

Precision = 89%

Berdasarkan rumusan di atas, dihasilkan nilai akurasi sebesar 85%, precision 78% dan recall 89%. Sedangkan skor data latih yang dihasilkan adalah 87% dan skor uji yang dihasilkan sebesar 85%. Berdasarkan hasil tersebut dapat disimpulkan bahwa model yang dihasilkan pada penelitian ini adalah fit.

Deployment Model

Setelah melalui beberapa tahapan preprocessing dan processing, selanjutnya adalah melakukan model deployment. Peneliti menyimpan model dalam ekstensi pickle untuk selanjutnya dimasukkan ke dalam sistem. Pada pembuatan sistem berbasis web, peneliti menggunakan framework streamlit.

Prediksi Penyakit Jantung

Masukkan Umur
40

Masukkan jenis nyeri dada. jika ATA = 1, NAP = 2, ASY = 0, TA = 3
1

Masukkan gula darah saat beristirahat
0

Masukkan Detak Jantung Maksimum
172

Apakah Mengalami Angin Duduk? Masukkan 1 jika ya, Masukkan 0 jika tidak
0

Masukkan depresi ST
0

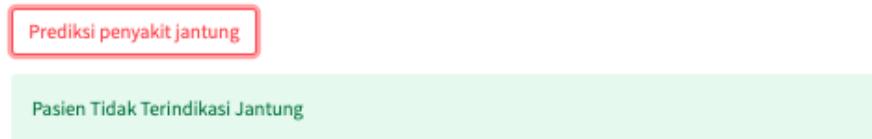
Masukkan Kemiringan segmen latihan puncak ST
2

Apakah pasien perempuan? Masukkan 1 jika ya, Masukkan 0 jika tidak
0

Apakah pasien laki-laki? Masukkan 1 jika ya, Masukkan 0 jika tidak
1

Prediksi penyakit jantung

Gambar 8. Uji Coba Sistem



Gambar 9 Hasil Uji Coba Sistem

Gambar 7 merupakan uji coba sistem hasil dari deployment model. Peneliti memasukkan inputan sesuai dengan indikator yang telah disediakan. Setelah memasukkan semua indikator, peneliti menekan tombol prediksi penyakit jantung sehingga tampil hasil klasifikasi yang terdapat pada gambar 8 yang menunjukkan bahwa pasien tidak terindikasi penyakit jantung berdasarkan indikator yang telah dimasukkan.

Pengujian Sistem Dengan Metode Black Box

Setelah melalui seluruh tahapan dari preprocessing hingga implementasi sistem, selanjutnya yaitu tahapan pengujian sistem. Hal ini bertujuan agar sistem dapat berfungsi dengan baik Ketika digunakan oleh user. Pada penelitian ini, peneliti menggunakan metode black box untuk melakukan pengujian. Seperti yang telah dijelaskan pada sub bab 2.7, pengujian dengan metode black box mengamati hasil input dan output tanpa mengetahui kode sumbernya. Berikut merupakan hasil pengujian sistem dengan metode black box.

Tabel 6 Hasil Pengujian Sistem

Aktivitas pengujian	Realisasi yang diharapkan	Hasil pengujian	kesimpulan
Menjalankan sistem	Tampil halaman utama beserta indikator dan tombol untuk memprediksi	Indikator yang ditampilkan sesuai dengan sistem yang telah dibuat	diterima

Memasukkan value sesuai dengan indikator yang telah disediakan	User dapat memasukkan hasil pemeriksaan ke dalam indikator yang telah disediakan	text box dapat berfungsi sesuai dengan yang diharapkan	diterima
Menekan tombol “prediksi penyakit jantung”	Sistem dapat melakukan prediksi sesuai indikator yang telah diinputkan	Sistem berfungsi dengan baik sehingga dapat menampilkan hasil prediksi	diterima

berdasarkan hasil pengujian sistem yang terdapat pada tabel 6 di atas, dapat dilihat bahwa semua pengujian menghasilkan hasil pengujian sesuai dengan realisasi yang diharapkan. Dimana sistem dapat menerima input dan menghasilkan output sesuai dengan yang diharapkan..

IV. SIMPULAN

Penelitian ini telah berhasil melakukan klasifikasi mulai dari preprocessing sampai evaluasi dan deployment model. Berdasarkan penelitian yang telah dilakukan, sistem dapat melakukan klasifikasi penyakit jantung dengan skor akurasi sebesar 85% dengan hyperparameter yang dihasilkan yaitu gamma sebesar 0.01, Const sebesar 0.1, dan kernel yang digunakan adalah polynomial karena bentuk dari dataset tidak linear. Adapun precision yang dihasilkan sebesar 78%, dan recall sebesar 89%. Skor latih yang dihasilkan pada penelitian ini adalah 85% dan skor ujinya adalah 87% sehingga jika dilihat dari gap antara skor latih dan skor uji, model tidak mengalami overfitting maupun underfitting. Dengan kata lain penelitian ini menghasilkan model yang fit. Sedangkan pada pengujian sistem, berdasarkan pengujian blacbox pada sistem berbasis web mendapatkan hasil pengujian sesuai dengan realisasi yang diharapkan.

UCAPAN TERIMA KASIH

Puji dan syukur kami panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, kami dapat menyelesaikan penelitian ini. Peneliti juga mengucapkan terima kasih kepada prodi Informatika Universitas Muhammadiyah Sidoarjo yang telah memberikan dukungan maupun bimbingan selama penelitian ini dilakukan.

REFERENSI

- [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*, vol. 4, no. 2, 2020, doi: 10.30865/mib.v4i2.2080.
- [2] S. Agustin, “Ciri-Ciri Sakit Jantung yang Harus Diwaspadai,” <https://www.alodokter.com/kenali-ciri-ciri-sakit-jantung>, May 08, 2022.
- [3] D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 2, p. 437, Apr. 2020, doi: 10.30865/mib.v4i2.2080.
- [4] Samsudiney, “Penjelasan Sederhana tentang Apa Itu SVM?,” <https://medium.com/@samsudiney/penjelasan-sederhana-tentang-apa-itu-svm-149fec72bd02>, Jul. 25, 2019.
- [5] A. F. HIDAYATULLAH, “MEMBUAT APLIKASI WEB SAINS DATA DENGAN MUDAH MENGGUNAKAN STREAMLIT,” <https://informatics.uii.ac.id/2021/03/15/streamlit-membuat-aplikasi-web-sains-data/>, Mar. 15, 2021.
- [6] M. A. Bianto, K. Kusrini, and S. Sudarmawan, “Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes,” *Creative Information Technology Journal*, vol. 6, no. 1, 2020, doi: 10.24076/citec.2019v6i1.231.
- [7] D. P. Utomo, P. Sirait, and R. Yunis, “Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5.0,” *Jurnal Media Informatika Budidarma*, vol. 4, no. 4, 2020.
- [8] A. Nurmasani and Y. Pristyanto, “ALGORITME STACKING UNTUK KLASIFIKASI PENYAKIT JANTUNG PADA DATASET IMBALANCED CLASS,” 2021. [Online]. Available: www.ejournal.unib.ac.id/index.php/pseudocode
- [9] A. Novita Sari and S. Alfionita, “Klasifikasi Penyakit Jantung Menggunakan Metode Naïve Bayes,” vol. 1, no. 1, pp. 22–26, 2022, doi: 10.12487/AMRI.v1i1.xxxxx.
- [10] P. R. Fitrayana and D. R. S. Saputro, “Algoritme Clustering Large Application (CLARA) untuk Menangani Data Outlier,” *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 5, pp. 721–725, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [11] P. R. , , D. A. Sihombing, S. Suryadiningrat, and Y. P. A. C. Yuda, . “Identifikasi Data Outlier (Pencilinan) dan Kenormalan Data Pada Data Univariat serta Alternatif Penyelesaiannya,” *Jurnal Ekonomi dan Statistik Indonesia*, pp. 307–316, 2022.
- [12] R. Fikri, “Handling Imbalanced Dataset,” <https://medium.com/@rusnandifikri96/handling-imbalanced-dataset-260378b2a21b>, Jul. 23, 2020.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.