



Plagiarism Checker X Originality Report

Similarity Found: 7%

Date: Thursday, April 20, 2023

Statistics: 261 words Plagiarized / 3618 Total words

Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

Sistem Prediksi Penyakit Jantung Berbasis Web **Menggunakan Metode SVM dan** Framework Streamlit Ary Putranto^{1,*}, Nuril Lutvi Azizah², Ika Ratna Indra Astutik³
^{1,2,3}Fakultas Sains dan Teknologi, Informatika, Universitas Muhammadiyah Sidoarjo, Sidoarjo, Indonesia Email: 1aryputranto.red@gmail.com, 2nurillutviazizah@umsida.ac.id, 3ikaratna@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 Abstract 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 bersifat 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 1.

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]. 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 bersifat 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. 2. Metode Penelitian 2.1 Tahapan Penelitian Tahapan penelitian merupakan gambaran umum terkait alur penelitian yang akan dilakukan dalam pengerjaan 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 2.2 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 Electrodigram istirahat _MaxHR _Detak jantung maksimum tercapai _ExerciseAngina _Induksi angina _Oldpeak _Depresi _HeartDisease _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. 2.3

Preprocessing 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. 2.3.1 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. 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. Tahap selanjutnya yaitu memfilter data yang nilai absolutnya kurang dari 3. 2.3.2 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. 2.3.3

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 sintesis". Pembuatan data training sintesis tersebut dibuat berdasarkan k-nearest neighbor [13]. 2.4 Processing Setelah data melalui tahapan preprocessing, selanjutnya adalah data dimasukkan ke tahap pemrosesan sehingga mendapatkan hasil yang diharapkan. 2.4.1 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 Membagi data menjadi dua bagian yaitu data latih dan data uji dengan persentase perbandingan 80% untuk data latih dan 20% untuk data uji. 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 _ _ 2.5 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]. 2.6 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.

/ Gambar 4 Desain Sistem Berbasis Web 2.7 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 kodenya. 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. 3. Hasil dan Pembahasan 3.1 Hasil Preprocessing 3.1.1 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 5 Hasil Filter Outlier 3.1.2

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 negative lemah dan 0.25 (korelasi positif lemah).

/ Gambar 6 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. 3.1.3 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 7 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. 3.2 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. 3.3 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 8 Hasil Confusion Matrix Berdasarkan hasil confusion matrix di atas, dapat dilakukan perhitungan precision, recall, dan accuracy. Berikut merupakan rumus beserta perhitungannya.

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{75}{75 + 9} = 0,853 \approx 85\% \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{75}{75 + 20} = 0,78 \approx 78\% \\ \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} = \frac{75 + 93}{75 + 9 + 20 + 93} = 0,89 \approx 89\% \end{aligned}$$

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. 3.4 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. / Gambar 9. Uji Coba Sistem / Gambar 10 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. 3.5 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. 4.

Kesimpulan 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. Daftar Pustaka [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.uui.ac.id/2021/03/15/streamlit-membuat-aplikasi-web-sains-data/>, Mar. 15, 2021. [6] M. A. Bianto, K. Kusriani, 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 (Pencilan) 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. [13] A. A. Arifiyanti and E. D. Wahyuni, "SMOTE: METODE PENYEIMBANG KELAS PADA KLASIFIKASI DATA MINING," ISSN 2686-6099, vol. 11, 2020. [14] A.

Liani, "Analisis Perbandingan Kernel Algoritma Support Vector Machine dalam Mengklasifikasikan Skripsi Teknik Informatika berdasarkan Abstrak," *JOINS (Journal of Information System)*, vol. 5, no. 2, 2020, doi: 10.33633/joins.v5i2.3715. [15] P. M. Kouate, "Machine Learning: GridSearchCV & RandomizedSearchCV," <https://towardsdatascience.com/machine-learning-gridsearchcv-randomizedsearchcv-d36b89231b10>, Sep. 11, 2020. [16] M. S. ANGGREANY, "Confusion Matrix," <https://socs.binus.ac.id/2020/11/01/confusion-matrix/>, Nov. 01, 2020. [17] R. Maulid, "Kriteria Jenis Teknik Analisis Data dalam Forecasting," <https://www.dqlab.id/kriteria-jenis-teknik-analisis-data-dalam-forecasting>, Jan. 14, 2022. [18] A. A. Suryanto, A. Muqtadir, and S.

Artikel, "PENERAPAN METODE MEAN ABSOLUTE ERROR (MEA) DALAM ALGORITMA REGRESI LINEAR UNTUK PREDIKSI PRODUKSI PADI Info Artikel : ABSTRAK," no. 1, p. 11,

2019. [19] Y. D. Wijaya and M. W. Astuti, "PENGUJIAN BLACKBOX SISTEM INFORMASI PENILAIAN KINERJA KARYAWAN PT INKA (PERSERO) BERBASIS EQUIVALENCE PARTITIONS," *Jurnal Digital Teknologi Informasi*, vol. 4, no. 1, 2021, doi: 10.32502/digital.v4i1.3163. [20] E. Agustin, A. Eviyanti, and N. Lutvi Azizah, "Deteksi Penyakit Epilepsi Melalui Sinyal EEG Menggunakan Metode DWT dan Extreme Gradient Boosting," vol. 7, no. 1, pp. 117–127, 2023, doi: 10.30865/mib.v7i1.5412.

INTERNET SOURCES:

<1% -
https://www.researchgate.net/publication/358561013_PREDIKSI_PENYAKIT_JANTUNG_CARDIOVASCULAR_MENGGUNAKAN_MODEL_ALGORITMA_KLASIFIKASI
<1% - <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>
<1% -
https://www.researchgate.net/publication/342236497_Anlisis_Komparasi_Metode_Klasifikasi_Data_Mining_dan_Reduksi_Atribut_Pada_Data_Set_Penyakit_Jantung
<1% -
<http://download.garuda.kemdikbud.go.id/article.php?article=1097407&val=16453&title=HUBUNGAN%20USIA%20JENIS%20KELAMIN%20DAN%20RIWAYAT%20PENYAKIT%20DENGAN%20KEJADIAN%20PENYAKIT%20JANTUNG%20KORONER%20DI%20INSTALASI%20GAWAT%20DARURAT%20RUMAHSAKIT%20TKII%20dr%20AK%20GANI%20PALEMBANG%20TAHUN%202016>
<1% -
<https://www.bunda.co.id/artikel/kesehatan/jantung/serangan-jantung-disebabkan-oleh-hal-tak-terduga-gejala-dan-faktor/>
<1% - <https://www.trivusi.web.id/2022/04/algoritma-svm.html>
<1% -
<https://cloudex.wg.ugm.ac.id/2020/01/21/implementasi-machine-learning-untuk-memprediksi-mood-pada-lagu/>
<1% - <https://jurnal.untan.ac.id/index.php/jepin/article/download/48053/75676591338>
<1% -
<https://adoc.pub/diagram-alir-penelitian-langkah-langkah-yang-dilakukan-dalam.html>
<1% - <https://repository.ittelkom-pwt.ac.id/7584/5/BAB%20III.pdf>
<1% -
https://www.researchgate.net/publication/338857001_Implementasi_Data_Mining_Untuk_Memprediksi_Penyakit_Jantung_Menggunakan_Metode_Naive_Bayes/fulltext/5e3029f8458515072d67fee0/Implementasi-Data-Mining-Untuk-Memprediksi-Penyakit-Jantung-Menggunakan-Metode-Naive-Bayes.pdf

1% - <https://journal.unnes.ac.id/sju/index.php/prisma/article/download/54585/21102/>

<1% -

<https://alamsyah599.medium.com/membuat-machine-learning-models-yang-dapat-mengatasi-outliers-regresi-robust-random-forest-1c95c9b698d4>

<1% -

<https://support.microsoft.com/id-id/office/apa-yang-harus-dilakukan-jika-kumpulan-data-terlalu-besar-untuk-kisi-excel-976e6a34-9756-48f4-828c-ca80b3d0e15c>

<1% - <https://accurate.id/akuntansi/koefisien-korelasi/>

<1% - <https://info.populix.co/articles/koefisien-korelasi-adalah/>

<1% - https://kc.umh.ac.id/17551/4/BAB_II.pdf

1% - <https://publikasi.dinus.ac.id/index.php/joins/article/download/3715/pdf>

<1% -

<https://ichi.pro/id/memahami-confusion-matrix-precision-recall-dan-f1-score-141157474959377>

<1% -

<https://www.mendeley.com/catalogue/e14b6c1e-6270-3b25-b377-48c30ba5b425/>

<1% - <https://bamai.uma.ac.id/2022/11/29/apa-itu-black-box-testing/>

<1% -

<https://lib.ui.ac.id/file?file=digital/126573-658.312%204%20NAD%20h%20-%20Hipotesis.pdf>

<1% - <https://algotech.netlify.app/blog/gridsearchcv/>

<1% - <https://eprints.umm.ac.id/63754/38/BAB%20II.pdf>

<1% - <https://eprints.uny.ac.id/13882/4/BAB%20III.pdf>

<1% - <https://www.dosenpendidikan.co.id/efektivitas-adalah/>

<1% -

<https://health.kompas.com/read/2020/08/03/103518868/10-penyebab-penyakit-jantung-yang-harus-diwaspadai>

<1% - <http://eprints.upj.ac.id/id/eprint/2730/14/14.%20Daftar%20Pustaka.pdf>

<1% - <http://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/2355>

<1% -

https://www.researchgate.net/publication/351170331_Algoritme_Stacking_Untuk_Klasifikasi_Penyakit_Jantung_Pada_Dataset_Imbalanced_Class

<1% -

<https://journal.unnes.ac.id/sju/index.php/prisma/citationstylelanguage/get/modern-language-association?submissionId=54585>

<1% - <https://dqlab.id/kriteria-jenis-teknik-analisis-data-dalam-forecasting>

<1% -

<https://123dok.com/document/qmvo4p8q-penerapan-metode-absolute-algoritma-regresi-linear-prediksi-produksi.html>