

Studi Komparatif Model Klasifikasi Kerentanan Penyakit Jantung Menggunakan Algoritma *Machine Learning*

Wiji Lestari¹, Sri Sumarlinda²

¹Fak. Ilmu Komputer, Universitas Duta Bangsa, wiji_lestari@udb.ac.id, Surakarta, Indonesia

²Fak. Ilmu Komputer, Universitas Duta Bangsa, sri_sumarlinda@udb.ac.id, Surakarta, Indonesia

Informasi Makalah

Submit : November 14, 2022
Revisi : Desember 9, 2022
Diterima : Juni 13, 2023

Kata Kunci :

Model klasifikasi
Analisis komparatif
Penyakit jantung
Machine learning

Abstrak

Penyakit jantung merupakan salah satu penyebab kematian baik di dunia maupun Indonesia. Perhatian awal dari penyakit jantung akan memudahkan pencegahan dan penyembuhannya. Tujuan penelitian ini adalah melakukan analisis komparatif model klasifikasi dengan berbagai algoritma machine learning untuk kerentanan penyakit jantung. Dataset diambil dari UCI machine Learning Respiratory dengan 300 data training dan 100 data testing. Parameter klasifikasi terdiri dari age, sex, systolic blood pressure, cholesterol, thalach, oldpeak dan slope, serta labelnya cardio. Model klasifikasi dibangun dengan algoritma Naïve Bayes, K-Nearest Neighbor (KNN), Decision Tree, random Forest, Backpropagation, Logistic Regression dan Support Vector machine (SVM). Hasil model klasifikasi dari pengukuran accuracy didapatkan Naïve Bayes (79,00%), KNN (63,00%), Decision Tree (66,00%), Random Forest (77,00%), Backpropagation (80,00%), Logistic Regression (81,00%) dan SVM (80,00%). Dari analisis komparatif pengukuran parameter accuracy, precision, recall dan F1 score maka model klasifikasi dengan algoritma Logistic Regression dan backpropagation menghasilkan performa terbaik.

Abstract

The heart disease is one of the leading causes of death both in the world and in Indonesia. Early attention to heart disease will facilitate its prevention and cure. The purpose of this study is to perform a comparative analysis of the classification model with various machine learning algorithms for heart disease susceptibility. The dataset was taken from UCI Machine Learning Respiratory with 300 training data and 100 testing data. The classification parameters consist of age, sex, systolic blood pressure, cholesterol, thalach, oldpeak and slope, and the label is cardio. The classification model was built using the Naïve Bayes algorithm, K-Nearest Neighbor (KNN), Decision Tree, random Forest, Backpropagation, Logistic Regression and Support Vector Machine (SVM). The results of the classification model from accuracy measurements obtained were Naïve Bayes (79.0%), KNN (63.00%), Decision Tree (66.00%), Random Forest (77.00%), Backpropagation (80.00%), Logistic Regression (81.00%) and SVM (80.00%). From the comparative analysis of parameter measurement accuracy, precision, recall and F1 score, the classification model with Logistic Regression and backpropagation algorithms produces the best performance.

1. Pendahuluan

Penyakit jantung menjadi salah satu penyebab utama kematian baik di Indonesia maupun di dunia. Organisasi Kesehatan dunia WHO (*World Health Organization*) pada tahun 2019 melaporkan kematian yang diakibatkan oleh penyakit jantung adalah sekitar 31% dari kematian di seluruh dunia (WHO, 2019). Di Indonesia berdasarkan laporan kementerian kesehatan pada tahun 2014 sekitar 35 % kematian disebabkan penyakit jantung (Wiharto et al., 2017). Penyakit jantung berkaitan erat dengan factor-faktor resiko seperti hipertensi, kolesterol dalam darah, gula darah, asam urat, umur, *life style* dan lain-lain. Pendeteksian dini penyakit jantung akan memudahkan penyembuhan dan mencegah akibat yang lebih fatal. Identifikasi kerentanan penyakit jantung merupakan cara awal untuk pencegahan terjadinya penyakit jantung yang lebih fatal. Identifikasi kerentanan penyakit jantung dilaksanakan dengan berdasarkan data factor-faktor resiko seseorang terhadap penyakit jantung. Kebanyakan masyarakat enggan memeriksakan diri ke dokter atau klinik, sehingga kadang penyakit jantung sudah akut baru memeriksakan diri. Perkembangan teknologi digital memudahkan kita untuk pengecekan data berkaitan dengan factor resiko penyakit jantung seperti hipertensi, kolesterol, gula darah dan sebagainya dengan alat-lat digital yang bisa dioperasikan sendiri (Lestari & Sumarlinda, 2022), (Sumarlinda & Lestari, 2022).

Perkembangan digitalisasi dalam kehidupan sehari-hari semakin masif. Muncul banyak metode, model dan algoritma dalam bidang komputasi dan teknologi informasi yang dapat memudahkan atau membantu pekerjaan manusia. *Data mining* dan *machine learning* adalah metode yang banyak digunakan pengolahan data di berbagai bidang kehidupan. Klasifikasi, *clustering*, prediksi, peramalan dan asosiasi merupakan fungsi-fungsi pada data mining. Fungsi klasifikasi merupakan fungsi data mining yang akan mengelompokkan sekelompok data dalam label-label tertentu (Goel et al., 2019). Proses klasifikasi dari data input menuju ke label-label yang telah ditentukan dapat menggunakan algoritma *machine learning*. Data input dalam proses klasifikasi dibuat ke

dalam *data training* dan *data testing*. Penerapan fungsi klasifikasi cukup banyak dan hamper di segala bidang. Fungsi klasifikasi digunakan untuk prediksi tingkat kepuasan pelanggan (Sains et al., 2020). Klasifikasi telah banyak diimplementasikan dalam bidang kesehatan, seperti klasifikasi tekanan darah menggunakan metode modular *neural network* (Pulido et al., 2019). Implementasi klasifikasi dengan *machine learning* untuk diagnosis kanker (Sidey-gibbons & Sidey-gibbons, 2019). Beberapa algoritma *machine learning* digunakan untuk klasifikasi pasien diabetes (Shang et al., 2021). Selain itu algoritma *machine learning* dan data mining juga dikembangkan dalam bentuk system pakar untuk deteksi penyakit gigi dengan metode Bayes (Hadi & Diana, 2019), *machine learning* juga digunakan untuk deteksi awal penyakit tulang belakang (Hayat & Latuny, 2020)

Model klasifikasi kerentanan penyakit jantung dapat digunakan untuk deteksi dini penyakit jantung pada seseorang. Model klasifikasi bergantung dari performa proses klasifikasinya dan evaluasi hasil klasifikasi. *Machine learning* dapat digunakan untuk model klasifikasi kerentanan penyakit jantung. *Data input* merupakan nilai dari factor-faktor resiko penyakit jantung. Penelitian-penelitian yang berkaitan dengan analisis komparatif model klasifikasi penyakit jantung sudah banyak dilaksanakan. Analisis komparatif klasifikasi penyakit jantung menggunakan algoritma *machine learning* dan *deep learning*, analisis ini menggunakan data inputnya, metode algoritmanya dan metode evaluasinya (Swathy & Saruladha, 2022). Model klasifikasi dibangun dengan KNN dan *Random Forest* dengan metode komparasi menggunakan PCA dan tidak menggunakan PCA (Alim et al., 2022). Model komparatif klasifikasi penyakit jantung dengan data HER (*electronic heart record*) dengan menggunakan metode *machine learning* dan *deep learning* dengan menghasikan *accuracy* terbaik untuk *Logistic Regression* dan SVM (Prusty et al., 2022). Analisis komparatif model klasifikasi penyakit jantung dengan menggunakan metode RIPPER, C4.5, MLP dan SVM. Dari hasil *accuracy* SVM menunjukkan nilai tertinggi (Khan et al., 2017). Analisis komparatif metode klasifikasi penyakit

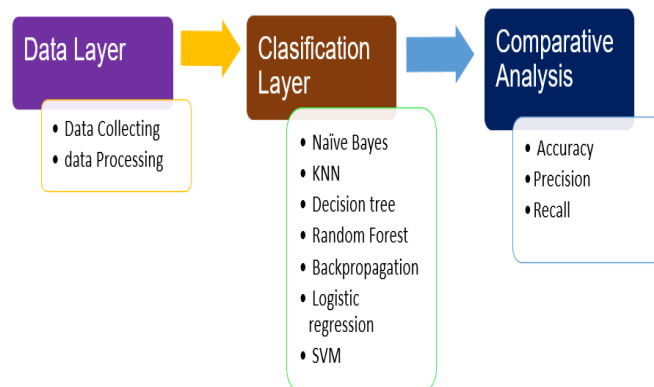
jantung dengan *Decision Tree*, *Support Vector Classifier*, *Random Forest*, *KNN* dan *Naïve Bayes*. Hasil analisis komparatif metode dengan *Naïve Bayes* memiliki performa terbaik (Puskhla et al., 2019). Komparatif model klasifikasi penyakit jantung dengan 6 algoritma *Decision Tree*, *Random Forest*, *SVM*, *KNN*, *Logistic Regression* dan *Gaussian Naïve Bayes*. Analisis ini membandingkan juga penggunaan *feature selection* dan tidak menggunakan *feature selection* pada proses klasifikasi (Dissanayake & Johar, 2021). Model yang hampir sama juga telah dikembangkan dengan 5 algoritma *machine learning* untuk klasifikasi penyakit jantung dengan ditambahi proses regresi dan normalisasi (Hasan, 2021).

Penelitian ini membangun model analisis komparatif metode klasifikasi kerentanan penyakit jantung berbasis *data mining* dan *mesin learning*. Pengembangan model klasifikasi penyakit jantung akan memudahkan deteksi dini penyakit jantung sehingga memudahkan ketatalaksanaan pencegahan dan penyembuhan. Model klasifikasi juga berkembang dan sangat variatif. Perbaikan model dapat dari pemrosesan *data input*, perbaikan dan keragaman algoritma dan pengujian performa proses seperti *accuracy*, *precision*, *recall* dan lain-lain. Pada penelitian ini komparatif model untuk membandingkan 7 algoritma *machine learning* yang digunakan untuk klasifikasi kerentanan penyakit jantung. Klasifikasi kerentanan penyakit jantung dibangun dengan 7 algoritma *machine learning* yaitu *Naïve Bayes*, *K-Nearest Neighbor* (KNN), *Decision Tree*, *Random Forest*, *Backpropagation*, *Logistic Regression* dan *Support Vector Machine* (SVM). Metode penelitian yang digunakan adalah penelitian deskriptik. Hasil proses klasifikasi akan dibandingkan dalam hal *accuracy*, *precision* dan *recall*.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan penelitian deskriptif kuantitatif. Penelitian ini mendeskripsikan proses klasifikasi kerentanan penyakit jantung dibangun dengan 7 algoritma *machine learning* yaitu *Naïve Bayes*, *K-Nearest Neighbor* (KNN), *Decision Tree*, *Random Forest*, *Backpropagation*, *Logistic Regression* dan *Support Vector*

Machine (SVM), yang hasil performanya akan dikomparasikan. Tahapan penelitian terlihat seperti gambar 1. di bawah ini:



Gambar 1. Tahapan Metode Penelitian

2.1 Data Collecting

Data input merupakan dataset pasien penyakit jantung yang diunduh dari UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/datasets/heart+disease>. Data input terdiri dari data pasien dengan 8 variabel yang berkaitan dengan factor resiko penyakit jantung, yaitu :

- **Age** – umur pasien (tahun, numerik)
- **Sex** – jenis kelamin pasien
 - a. 0 wanita
 - b. 1 pria
- **Systolic Blood Pressure** – tekanan darah (mmHg, numerik)
- **Cholesterol** – serum kolesterol (mg/dl, numerik)
- **Thalach** – Nilai maksimum detak jantung (numerik)
- **Oldpeak** – depresi ST yang diakibatkan kegiatan fisik relative terhadap saat istirahat (numerik)
- **Slope** – Kemiringan pincak ST segmen (numerik)
 - a. 1 upsloping
 - b. 2 flat
 - c. 3 downsloping
- **Cardio** – diagnosis penyakit jantung (nominal)
 - a. 0 terindikasi sehat
 - b. 1 terindikasi penyakit jantung

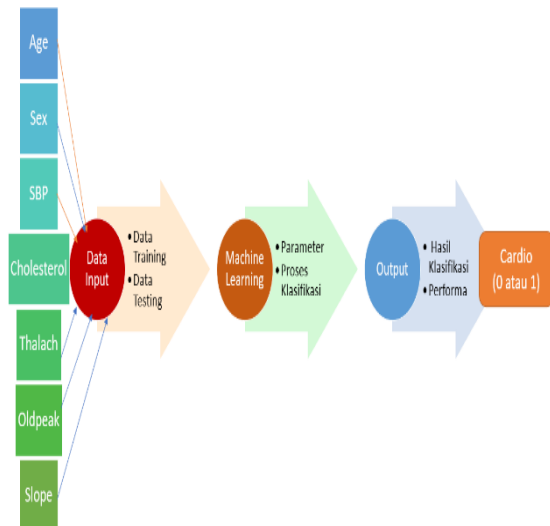
2.2 Data Processing

Dataset yang dijadikan data input sebelum dilaksanakan proses klasifikasi maka

dilakukan *data processing* terlebih dahulu. *Data processing* meliputi penghilangan *missing data* dan juga data yang tidak lengkap. Setelah melalui proses ini didapatkan data input sebanyak 400. Dari 400 data pasien penyakit jantung ini diseparasi menjadi *data training* sebanyak 300 dan *data testing* sebanyak 100. *Data training* dan *data testing* selanjutnya akan diproses dengan fungsi klasifikasi menggunakan 7 algoritma *machine learning*.

2.3 Metode Klasifikasi

Proses klasifikasi yaitu proses pengelompokan data input berdasarkan variabel faktor-faktor resiko penyakit jantung ke label-label yang ditentukan dari awal. Pada penelitian ini data output berkaitan dengan variabel **cardio**, bernilai 1 jika rentan terhadap penyakit jantung dan 0 berarti sehat. Ada 7 variabel untuk proses klasifikasi yaitu : age, sex, systolic blood pressure, kolesterol, thalach, oldpeak dan slope. Algoritma machine learning digunakan untuk proses klasifikasi. Proses ini terlihat seperti gambar 2. di bawah :



Gambar 2. Proses Klasifikasi

Pengembangan model klasifikasi menggunakan 7 algoritma machine learning yaitu *Naïve Bayes*, *K-Nearest Neighbor (KNN)*, *Decision Tree*, *Random Forest*, *Backpropagation*, *Logistic Regression* dan *Support Vector Machine (SVM)*. Implementasi program menggunakan RapidMiner Studio. Gambaran singkat dari

algoritma machine learning terpakai seperti pada tabel 1. di bawah ini :

Tabel 1. Algoritma *Machine Learning* pada Model Klasifikasi

No	Algoritma	Keterangan/parameter
1	Naïve Bayes	Laplace correction
2	K-Nearest Neighbor (KNN)	K = 3, jarak Euclidian
3	Decision Tree	Information Gain, maximum depth = 5.
4	Random Forest	Jumlah trees = 10, gain ratio, maximum depth = 10
5	Bcakpropagation	Epochs = 500, learning rate = 0,01, fungsi aktivasi = sigmoid, pola jaringan 7-2-2
6	Logistic Regression	Solver = auto, standard
7	Support Vector Machine (SVM)	Tipe kernel = dot, kernel cache = 200, jumlah epochs = 500

Decision tree learning adalah salah satu metode learning yang berusaha menemukan fungsi-fungsi pendekatan yang bersifat diskrit, bisa menangani missing data dan mampu mempelajari ekspresi-ekspresi *disjunctive* (ekspresi OR). Algoritma *Decision tree learning* termasuk ke dalam penerapn data mining untuk fungsi prediksi dan klasifikasi. *Decision tree learning* membangun pohon keputusan (*decision tree*) dari sebuah data training yng berupa *record-record* dalam basis data. *Decision tree learning* terdiri dari kumpulan simpul (node) yang dihubungkan oleh cabang cabang tersebut bergerak ke bawah dari root node dan berakhir pada leaf node. Leaf node adalah node yang tidak dapat dipecah lagi dan mrepresentasikan prediksi jawaban dari masalah.

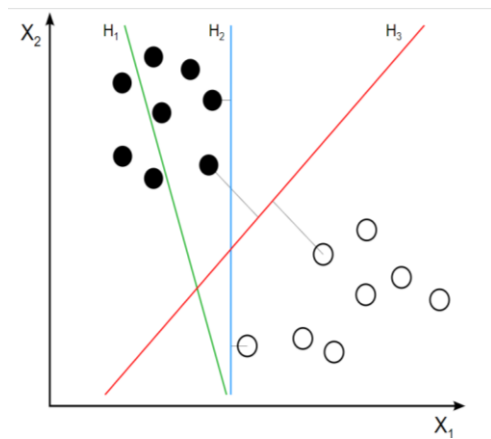
Bayesian algorithm atau juga disebut *Naïve Bayes algorithm* adalah salah satu algoritma klasifikasi berdasarkan teorema Bayesian. Bayesian algorithm dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Bayesian algorithm dapat digunakan untuk data input bertipe nominal maupun numerik.

Algoritma *K-nearest neighbor*, juga dikenal sebagai *KNN* atau *k-NN*, adalah *classifier* pembelajaran non-parametrik yang diawasi, yang menggunakan kedekatan untuk

membuat klasifikasi atau prediksi tentang pengelompokan titik data individual. Meskipun dapat digunakan untuk masalah regresi atau klasifikasi, ini biasanya digunakan sebagai algoritme klasifikasi, bekerja dengan asumsi bahwa titik serupa dapat ditemukan di dekat satu sama lain.

Random forest adalah metode pembelajaran ansambel untuk klasifikasi, regresi, dan tugas lain yang beroperasi dengan membangun banyak pohon keputusan pada waktu pelatihan. Untuk tugas klasifikasi, hasil random forest adalah kelas yang dipilih oleh sebagian besar pohon. Untuk tugas regresi, prediksi rata-rata atau rata-rata dari masing-masing pohon dikembalikan keputusan acak yang benar untuk kebiasaan overfitting pohon keputusan ke set pelatihan mereka.

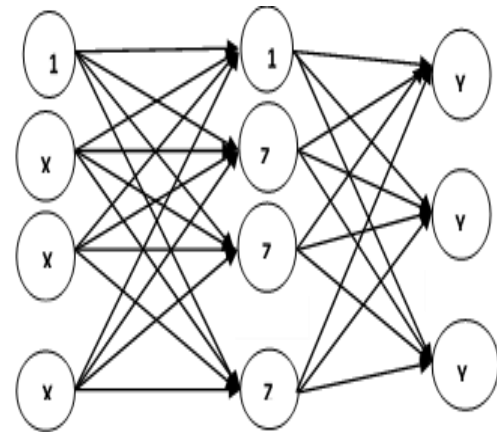
Support vector machine (SVM) adalah sekumpulan metode pembelajaran terawasi untuk klasifikasi, regresi, dan deteksi outlier. Keunggulan SVM adalah efektif untuk ruang berdimensi tinggi dan menggunakan subset titik pelatihan pada fungsi keputusan (vektor pendukung). SVM membentuk *hyperplane* dalam ruang dimensi tak terbatas yang dapat digunakan untuk regresi dan klasifikasi.



Gambar 3. *Support Vector Machine*

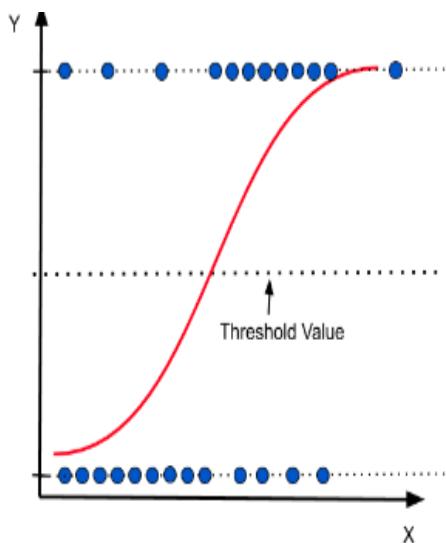
Algoritma *Backpropagation* atau yang dalam bahasa Indonesia ada yang menerjemahkan dengan Propagasi Balik, dirumuskan pertama kali oleh Werbos dan dipopulerkan oleh Rumelhart dan cClelland. Algoritma ini menggunakan metode pembelajaran Supervised Learning. Algoritma ini dikatakan Propagasi Balik/*Backpropagation* karena ketika jaringan diberikan pola masukan sebagai pola pelatihan maka pola tersebut menuju ke unit-

unit pada lapisan keluaran. Kemudian unit-unit lapisan keluaran memberikan tanggapan yang disebut sebagai keluaran jaringan. Saat keluaran jaringan tidak sama dengan keluaran keluaran yang diharapkan maka keluaran akan menyebar mundur (*backward*) pada lapisan tersembunyi diteruskan ke unit pada lapisan masukan.



Gambar 4. *Backpropagation*

Logistic regression adalah proses pemodelan probabilitas hasil diskrit yang diberikan variabel input. Model *logistic regression* yang paling umum adalah hasil biner; sesuatu yang dapat mengambil dua nilai seperti benar/salah, ya/tidak, dan sebagainya. Regresi logistik multinomial dapat memodelkan skenario di mana terdapat lebih dari dua kemungkinan hasil yang berlainan. *Logistic regression* adalah metode analisis yang berguna untuk masalah klasifikasi, di mana Anda mencoba untuk menentukan apakah sampel baru paling cocok dengan suatu kategori. Karena aspek keamanan dunia maya merupakan masalah klasifikasi, seperti deteksi serangan, regresi logistik merupakan teknik analitik yang berguna.



Gambar 5. Logistic Regression

Hasil proses klasifikasi dari 7 variabel akan dipetakan ke dalam 2 label yaitu rentan penyakit jantung (1) dan tidak rentan (0). Proses klasifikasi data dipengaruhi oleh performa masing-masing algoritma dan juga parameter yang dipakai.

2.4 Analisis Komparatif

Hasil proses klasifikasi akan dikomparasi antar algoritma dengan menggunakan pengukuran matrik performa. Pengukuran performa model klasifikasi meliputi *precision*, *recall*, *accuracy* dan *F1 score*. Uraian tentang parameter performa sebagai berikut:

➤ Precision (P)

$$P = \frac{TP}{TP+FP} \quad (1)$$

➤ Recall (R)

$$R = \frac{TP}{TP+FN} \quad (2)$$

➤ Accuracy

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

➤ F1 Score

$$F1\ Score = 2 \times \frac{P \times R}{P+R} \quad (4)$$

Keterangan :

TP (*true positive*) = pasien berpenyakit jantung dan hasil klasifikasi juga berpenyakit jantung

FP (*false positive*) = pasien tidak berpenyakit jantung, tetapi hasil klasifikasi berpenyakit jantung

TN (*true negative*) = pasien tidak berpenyakit jantung dan hasil klasifikasi juga tidak berpenyakit jantung

FN (*false negative*) = pasien berpenyakit jantung tetapi hasil klasifikasi tidak berpenyakit jantung

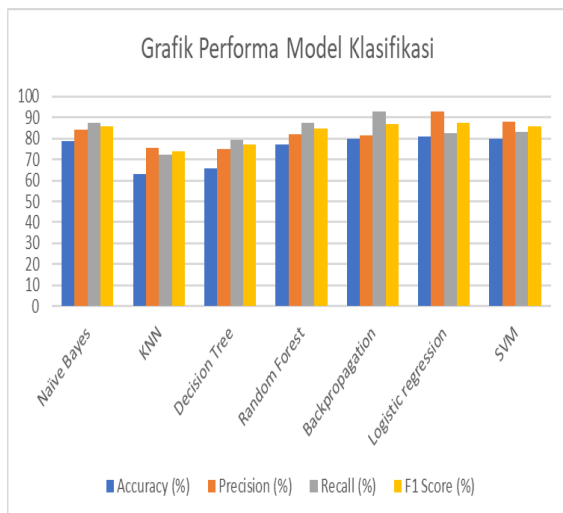
3. Hasil dan Pembahasan

Model klasifikasi untuk kerentanan penyakit jantung menggunakan 300 *data training* dan 100 *data testing*. Data input menggunakan 7 variabel untuk parameter klasifikasi dan 1 parameter untuk label. Proses klasifikasi 7 algoritma machine learning yaitu *Naïve Bayes*, *K-Nearest Neighbor* (KNN), *Decision Tree*, *Random Forest*, *Backpropagation*, *Logistic Regression* dan *Support Vector Machine* (SVM). Hasil performa klasifikasi dari berbagai model ini adalah seperti pada tabel 2. di bawah ini:

Tabel 2. Komparasi Performa Model Klasifikasi

No	Algoritma	Accuracy (%)	Precision (%)	Recall (%)	F1
					Score (%)
1	Naïve Bayes	79,00	84,00	87,50	85,71
2	KNN	63,00	75,36	72,22	73,76
3	Decision Tree	66,00	75,00	79,17	77,03
4	Random Forest	77,00	81,82	87,50	84,56
5	Backpropagation	80,00	81,71	93,06	87,02
6	Logistic regression	81,00	93,06	82,72	87,59
7	SVM	80,00	88,24	83,33	85,71

Secara grafik komparasi performa hasil klasifikasi dari berbagai metode adalah seperti pada gambar 3. di bawah ini:



Gambar 3. Komparasi Performasi Model Klasifikasi

Berdasarkan tabel 2. dan gambar 3. terlihat untuk *accuracy* algoritma *Logistic Regression* mempunyai nilai tertinggi 81,00% diikuti *Backpropagation* dan *Support Vector machine* (SVM) dengan nilai 80,00%. Untuk nilai terendah *accuracy* adalah *K-Nearest Neighbor* dengan nilai 63,00%. Untuk *precision* tertinggi algoritma *Logistic Regression* dengan nilai 93,06% sedangkan nilai terendah untuk algoritma *Decision Tree* dengan nilai 75,00%. Parameter *recall* nilai tertinggi algoritma *Backpropagation* dengan nilai 93,06% dan nilai terendah algoritma *KNN* dengan nilai 72,22%. Pengukuran *F1 score* nilai tertinggi algoritma *Logistic Regression* dengan nilai 87,59% dan *Backpropagation* 87,02%, sedangkan nilai terendah algoritma *KNN* dengan nilai 73,76%. Dari berbagai parameter terlihat bahwa performa klasifikasin untuk algoritma *Logistic Regression* dan *Backpropagation* memiliki performa yang lebih baik. Kedua algoritma baik *Logistic regression* dan *Backpropagation* menggunakan fungsi *sigmoid* untuk proses klasifikasinya. Untuk *Logistic Regression* merupakan pengembangan dari *Linear regression* .

$$Y_{in} = b_0 + b_1 \cdot X \quad (5)$$

$$Y_{out} = F(X) = \frac{1}{1 + e^{b_0 + b_1 X}} \quad (6)$$

Untuk algoritma *Backpropagation* fungsi *sigmoid* digunakan untuk fungsi aktivasi :

$$Y_{in} = \sum_i^n w_{ij} \cdot X_i + b \quad (7)$$

$$Y_{out} = \frac{1}{1 + e^{\sum_i^n w_{ij} \cdot X_i + b}} \quad (8)$$

Fungsi *sigmoid* pada kedua algoritma di atas dapat menangani *outlier data* dengan baik dibandingkan dengan fungsi lain, misalnya fungsi linier.

Logistic Regression adalah algoritme klasifikasi biner ML terawasi yang banyak digunakan di sebagian besar aplikasi. Ini bekerja pada variabel dependen kategorikal yang hasilnya dapat berupa variabel kategorikal diskrit atau biner 0 atau 1. Fungsi *sigmoid* digunakan sebagai fungsi biaya. Fungsi *sigmoid* memetakan nilai real yang diprediksi ke nilai probabilistik antara '0' dan '1'. Untuk memprediksi model ML regresi logistik penyakit jantung digunakan, pertama model LR dilatih dengan lima kondisi pemisahan dan diuji dengan data uji untuk prediksi untuk mendapatkan akurasi terbaik dan untuk menemukan perilaku model. Algoritma menghasilkan kategori 1 dan 0 untuk ada dan tidak adanya penyakit jantung.

4. Simpulan

Model klasifikasi kerentanan penyakit jantung dengan *dataset* 400 data pasien, yang dibagi menjadi 300 *data training* dan 100 *data testing*. Proses klasifikasi menggunakan 7 algoritma machine learning yaitu *Naive Bayes*, *K-Nearest Neighbor* (KNN), *Decision Tree*, *Random Forest*, *Backpropagation*, *Logistic Regression* dan *Support Vector Machine* (SVM). Analisis komparatif model klasifikasi menggunakan *accuracy*, *precision*, *recall* dan *F1 score*. parameter *accuracy* algoritma *Logistic Regression* mempunyai nilai tertinggi 81,00% diikuti *Backpropagation* dan *Support Vector machine* (SVM) dengan nilai 80,00%. Untuk *precision* tertinggi algoritma *Logistic Regression* dengan nilai 93,06%. Parameter *recall* nilai tertinggi algoritma *Backpropagation* dengan nilai 93,06%.

Pengukuran *F1 score* nilai tertinggi algoritma *Logistic Regression* dengan nilai 87,59% dan *Backpropagation* 87,02%. Dari analisis komparatif beberapa metrik performa model klasifikasi maka dapat disimpulkan bahwa algoritma *Logistic regression* dan *Backpropagation* mempunyai performa lebih baik.

5. Referensi

- Alim, M. A., Haris, A., Haider, W., & Masroor, A. (2022). A Comparative Study of Heart Disease Prediction Based on Principal Component Analysis And Classification Techniques. *Webology*, 19(3), 11–16. <https://doi.org/10.1109/DASA54658.2022.9765241>
- Dissanayake, K., & Johar, M. G. M. (2021). Comparative study on heart disease prediction using feature selection techniques on classification algorithms. *Applied Computational Intelligence and Soft Computing*, 2021. <https://doi.org/10.1155/2021/5581806>
- Goel, S., Deep, A., Srivastava, S., & Tripathi, A. (2019). Comparative Analysis of various Techniques for Heart Disease Prediction. *2019 4th International Conference on Information Systems and Computer Networks, ISCON 2019*, 88–94. <https://doi.org/10.1109/ISCON47742.2019.9036290>
- Hadi, F., & Diana, Y. (2019). Diagnosa Penyakit Gigi dengan Metode Bayes. *SATIN – Sains Dan Teknologi Informasi Sistem Pakar*, 5(2).
- Hasan, R. (2021). Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction. *ITM Web of Conferences*, 40. <https://doi.org/doi.org/10.1051/itmconf/20214003007>
- Hayat, C., & Latuny, A. A. (2020). Rancang Bangun Aplikasi Informasi Awal Penyakit Tulang Belakang dengan Metode Forward Chaining. *SATIN – Sains Dan Teknologi Informasi*, 6(1).
- Khan, S. N., Nawari, N. M., Shahzad, A., Ullah, A., Mushtaq, M. F., Mir, J., & Aamir, M. (2017). Comparative analysis for heart disease prediction. *International Journal on Informatics Visualization*, 1(4–2), 227–231. <https://doi.org/10.30630/joiv.1.4-2.66>
- Lestari, W., & Sumarlinda, S. (2022). Implementation of K-Nearest Neighbor (Knn) and Support Vector Machine (Svm) for Classification Cardiovascular Disease. *International Journal of MultiSciences*, 2(10), 30–36. <https://archive.ics.uci.edu/ml/datasets/heart+disease>.
- Prusty, S., Patnaik, S., & Dash, S. K. (2022). Comparative analysis and prediction of coronary heart disease. *Indonesian Journal of Electrical Engineering and Computer Science*, 27(2), 944–953. <https://doi.org/10.11591/ijeecs.v27.i2.p944-953>
- Pulido, M., Melin, P., & Prado-Arechiga, G. (2019). Blood pressure classification using the method of the modular neural networks. *International Journal of Hypertension*, 2019. <https://doi.org/10.1155/2019/7320365>
- Puskhla, V., Agaly, T., & Angayarkammii, S. A. (2019). Comparative Study of Heart Disease Prediction using Machine Learning Algorithms. *International Journal of Innovation in Engineering and Technology*, 12(4), 270–272. <https://doi.org/10.22214/ijraset.2022.44895>
- Sains, S., Ariani, F., & Taufik, A. (2020). Perbandingan Metode Klasifikasi Data Mining untuk Prediksi Tingkat Kepuasan Pelanggan Telkomsel Prabayar. *SATIN-Sains Dan Teknologi Informasi*, 6(2), 46–54. <https://doi.org/10.33372/stn.v6i2.666>
- Shang, Y., Jiang, K., Wang, L., Zhang, Z., Zhou, S., Liu, Y., Dong, J., & Wu, H. (2021). The 30-days hospital

- readmission risk in diabetic patients: predictive modeling with machine learning classifiers. *BMC Medical Informatics and Decision Making*, 21(Suppl 2), 1–11. <https://doi.org/10.1186/s12911-021-01423-y>
- Sidey-gibbons, J. A. M., & Sidey-gibbons, C. J. (2019). *Machine learning in medicine : a practical introduction*. 4, 1–18.
- Sumarlinda, S., & Lestari, W. (2022). Aplikasi K-Nearest Neighbor (KNN) untuk Klasifikasi Penyakit Kardiovaskuler. *Prosiding Seminar Nasional Teknologi ...*, 55, 259–261. <http://ojs.uib.ac.id/index.php/Senatib/article/download/1897/1487>
- Swathy, M., & Saruladha, K. (2022). A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques. *ICT Express*, 8(1), 109–116. <https://doi.org/10.1016/j.icte.2021.08.021>
- WHO. (2019). *WHO Health Statistics Overview 2019* (pp. 1–16).
- Wiharto, W., Kusnanto, H., & Herianto, H. (2017). Clinical Decision Support System for Assessment Coronary Heart Disease Based on Risk Factor. *Indian Journal of Science and Technology*, 10(22), 1–12. <https://doi.org/10.17485/ijst/2017/v10i22/84940>