

Komparasi Algoritma Supervised Learning dan Feature Selection pada Klasifikasi Penyakit Gagal Jantung

Khrisna Irham Fadhil Pratama¹, Kusnawi²

irhamfp@students.amikom.ac.id, khusnawi@amikom.ac.id

Universitas Amikom Yogyakarta

Informasi Artikel

Diterima : 30 Okt 2023

Direview : 2 Nov 2023

Disetujui : 20 Des 2023

Kata Kunci

Klasifikasi, Supervised Learning, Forward Selection, Pearson Corellation, Gagal Jantung

Abstrak

Penyakit gagal jantung merupakan penyakit yang mematikan yang ada di dunia, gagal jantung terjadi karena kondisi atau adanya kelainan otot-otot pada jantung. Pada tahun 2021, data yang ada pada WHO menunjukkan kematian dikarenakan penyakit jantung mencapai 17,8 juta jiwa. Salah satu cara yang dapat dilakukan yaitu dengan klasifikasi dengan menggunakan dataset public Kaggle. Penelitian ini bertujuan memkomparasi algoritma *supervised learning* dan metode *feature selection* yang terbaik, guna memperoleh hasil analisis data dengan akurasi yang baik dalam klasifikasi. Penerapan algoritma SVM, KNN, Naïve Bayes tanpa menggunakan *feature selection* algoritma SVM unggul menghasilkan *accuracy* 88.41%. Penerapan *forward selection* pada algoritma SVM, KNN, Naïve Bayes, algoritma SVM unggul dengan nilai *accuracy* 89.86%. Penerapan *Pearson corellation* pada algoritma SVM, KNN, Naïve Bayes, algoritma KNN unggul menghasilkan *accuracy* 90.58%. Penerapan *feature selection* baik *forward selection* dan *Pearson corellation* mampu meningkatkan performa akurasi, akan tetapi penerapan *Pearson corellation* pada penelitian ini lebih baik dalam meningkatkan akurasi.

Keywords

Classification, Supervised Learning, Forward Selection, Pearson Corellation, Heart Failure

Abstract

Heart failure is a deadly disease in the world, and heart failure occurs due to conditions or abnormalities in the muscles of the heart. In 2021, WHO data shows that deaths due to heart disease will reach 17.8 million people. One way that can be done is by classification using the public Kaggle dataset. This research aims to compare the best supervised learning algorithms and feature selection methods, in order to obtain data analysis results with good accuracy in classification. The application of the SVM, KNN, and Naive Bayes algorithm without using the superior SVM feature selection algorithm produces an accuracy of 88.41%. Application of forward selection to the SVM, KNN, and Naive Bayes algorithms, the SVM algorithm is superior, with an accuracy value of 89.86%. The application of Pearson correlation to the SVM, KNN, Naive Bayes algorithm, the superior KNN algorithm produces an accuracy of 90.58%. The application of feature selection, both forward selection and Pearson correlation, can improve accuracy performance; however, the application of Pearson correlation in this research is better in increasing accuracy.

A. Pendahuluan

Setiap manusia beraktifitas dengan menggunakan jantung. Jantung merupakan hal yang penting bagi manusia untuk melakukan aktifitas sehari-hari dan melakukan proses sirkulasi darah pada tubuh. Organ jantung memiliki salah satu penyakit yaitu penyakit gagal jantung. Gagal jantung dapat disebut juga dengan gagal jantung kongestif (*Congestive Heart Failure*). Penyakit gagal jantung merupakan penyakit yang mematikan yang ada di dunia, gagal jantung terjadi karena kondisi atau adanya kelainan otot-otot pada jantung[1]. Gagal jantung merupakan kondisi progresif sehingga jantung dalam memompa darah tidak mampu dalam memenuhi kebutuhan tubuh[2].

Pada tahun 2021 data yang ada pada WHO kematian dikarenakan penyakit jantung mencapai 17,8 juta jiwa[3]. Kematian penderita penyakit jantung 85% disebabkan oleh penyakit gagal jantung[4]. Penyakit gagal jantung yang berada di Indonesia pada tahun 2018 mencapai 1.017.290 penderita[5]. Penyakit gagal jantung diprediksi akan meningkat sampai 2030[4]. Berdasarkan jumlah penderita penyakit gagal jantung yang tinggi dan akan terus meningkat serta pentingnya organ jantung, memprediksi penyakit gagal jantung menjadi keutamaan seorang dokter dan tenaga medis[6]. Dengan begitu dibutuhkan sebuah metode yang efektif guna memprediksi penyakit gagal jantung.

Salah satu cara yang dapat dilakukan yaitu dengan cara klasifikasi penyakit gagal jantung. Klasifikasi merupakan metode pada data mining supervised learning dengan menggunakan data latih sebagai data pembelajaran dalam melakukan prediksi kelas pada sebuah objek[7]. Pratama dkk. [8] menerapkan algoritma K-Nearest Neighbor untuk mengklasifikasi penyakit gagal jantung. Data yang digunakan sebanyak 918 data dan 12 atribut. Klasifikasi tersebut memperoleh hasil akurasi sebesar 70,65%. Prabiantissa [9] menerapkan klasifikasi penyakit hati menggunakan algoritma Support Vektor Machine, K-NN, dan Naïve Bayes. Data yang digunakan dalam klasifikasi terdiri dari 14 atribut dan 303 instace. Akurasi yang dihasilkan pada klasifikasi penyakit hati dengan algoritma Naïve Bayes sebesar 82,42%, K-NN 68,13%, dan SVM sebesar 84,62%. Berdasarkan penelitian tersebut beberapa akurasi yang dihasilkan masih kurang bagus maka perlu adanya optimasi salah satunya dengan feature selection.

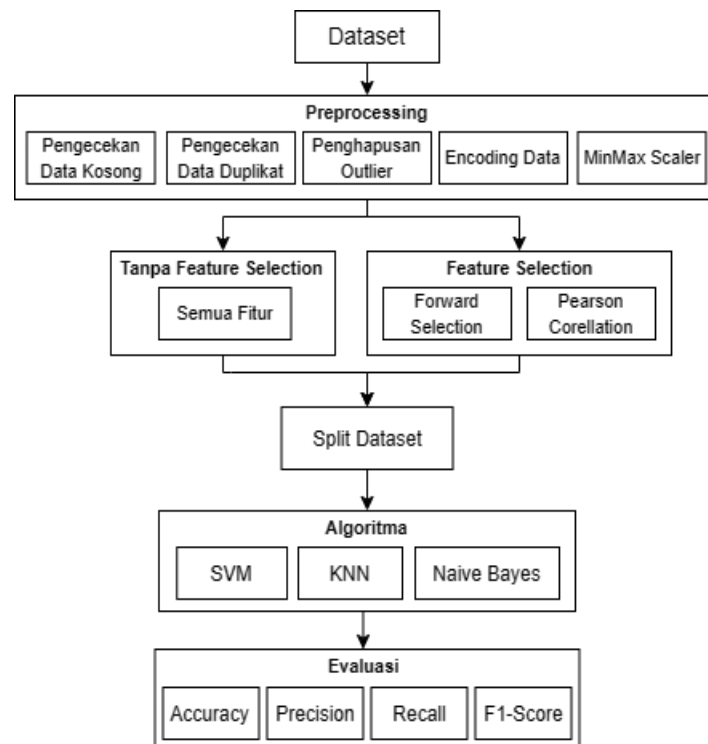
Feature selection merupakan sebuah proses identifikasi subset fitur yang paling relevan pada kumpulan fitur yang berguna untuk meningkatkan kinerja algoritma[10]. Penelitian yang dilakukan Abu Salam dkk dengan judul Implementasi Algoritma K-Nearest Neighbor Berbasis Forward Selection Untuk Prediksi Mahasiswa Non Aktif Universitas Dian Nuswantoro Semarang. Akurasi yang diperoleh menggunakan algoritma K-NN sebesar 96,43% menjadi 97,27% setelah menggunakan forward selection[11]. Adapun Nova Tri Romadloni dan Hilman F Pardede menerapkan seleksi fitur dengan Pearson correlation menggunakan algoritma Logistic Regresion, Naïve Bayes, dan SVM. Penelitian membandingkan akurasi sebelum dan sesudah menggunakan Pearson corellation. Akurasi sebelum menggunakan Pearson corellation sebesar 76,72% Logistic Regresor, 68,15% Naïve Bayes, 94,65% SVM sedangkan menggunakan Pearson corellation berdasarkan seleksi fitur akurasi menjadi 98,80% Logistic Regresor, 87,97% Naïve Bayes, 98,12% SVM [12]. Berdasarkan penelitian tersebut forward selection dan Pearson

corellation dapat meningkatkan kinerja algoritma dengan meningkatnya performa akurasi yang dihasilkan.

Penelitian ini menggunakan algoritma supervised learning yaitu SVM, K-NN, dan Naïve Bayes. Selain itu, penelitian ini menggunakan metode *feature selection* guna menentukan atribut yang terbaik pada data penyakit gagal jantung sehingga dapat melakukan peningkatan performa akurasi. *Feature selection* yang digunakan pada penelitian ini yaitu *forward selection* dan *Pearson corellation*. Penelitian ini bertujuan mengkomparasi algoritma *supervised learning* dan metode *feature selection* yang terbaik, guna memperoleh hasil analisis data dengan akurasi yang baik dalam melakukan klasifikasi penyakit gagal jantung. Penelitian ini diharapkan dapat membantu para dokter maupun tenaga medis dalam mengklasifikasi penyakit gagal jantung.

B. Metode Penelitian

Penelitian ini befokus pada pembangunan model klasifikasi penyakit gagal jantung dengan mengkomparasi performa yang dihasilkan dari algoritma SVM, KNN dan Naïve Bayes tanpa *feature selection* dan menggunakan *feature selection* yaitu *forward selection* dan *Pearson corellation*. Alur penelitian diawali dari pengumpulan data, preprocessing, penggunaan semua fitur tanpa *feature selection* dan penggunaan fitur yang telah terseleksi dengan *forward selection* maupun *Pearson corellation*. Selanjutnya yaitu split dataset dengan membagi data menjadi data training dan data testing. Setelah itu, masuk ke tahap pemodelan dan yang terakhir tahap evaluasi. Alur penelitian dapat digambarkan pada Gambar 1.



Gambar 1. Alur Penelitian

1. Dataset

Data yang digunakan pada penelitian menggunakan dataset public *Heart Failure Prediction Dataset* berasal dari Kaggle. Data berjumlah 918 data dengan 12 fitur. Data dimuat tahun 2021 dan dapat diunduh melalui <https://www.Kaggle.com/datasets/fedesoriano/heart-failure-prediction>.

2. Preprocessing

Preprocessing adalah teknik yang digunakan dalam mengubah data mentah menjadi lebih terstruktur yang berguna dalam pengolahan selanjutnya[13]. Pada penelitian ini proses preprocessing berupa pengecekan jumlah data kosong, pengecekan duplikasi data, pengecekan outlier dilakukan dengan perhitungan menggunakan *Interquartile Range* (IQR)[14]. Tahap preprocessing selanjutnya yaitu *encoding data* dan *MinMax Scaler*. *Encoding data* merupakan proses merubah data yang berbentuk kategorikal diubah menjadi bentuk numerik[15]. *MinMax Scaler* merupakan metode normalisasi dengan mengubah skala menjadi 0 dan 1 pada kumpulan data[14].

3. Forward Selection

Forward selection adalah salah satu metode wrapper yang digunakan untuk memilih fitur yang dilakukan sebelum masuk kedalam proses klasifikasi dengan pemilihan fitur yang relevan dengan data untuk mempengaruhi hasil klasifikasi dan meningkatkan kinerja pada metode klasifikasi[16]. *Forward selection* memilah fitur yang tidak digunakan pada saat itersi fitur dimulai, kemudian fitur akan ditambahkan dalam subset fitur yang terpilih sebelumnya[17].

4. Pearson Correlation

Pearson corellation adalah matrik statistik yang dapat menghitung kekuatan dan korelasi pada dua variabel yang dapat diterapkan pada analisis data, klasifikasi, clustering, analisis keuangan, dan lain-lain[18]. Korelasi dapat berupa korelasi positif dan negatif, korelasi positif adalah korelasi yang menghasilkan nilai bersifat searah sedangkan korelasi negatif merupakan korelasi yang menghasilkan nilai tidak searah. *Pearson corellation* dapat dilihat korelasi antar fitur melalui *heatmap* dengan memiliki fitur target yaitu fitur label[19].

5. Support Vector Machine

Support vector machine (SVM) merupakan metode *supervised learning* yang dapat mengatasi masalah yang berkaitan dengan klasifikasi dan regresi linier maupun non linear[20]. SVM memiliki konsep yang sederhana yaitu mencari hyperplane terbaik yang berguna untuk memisahkan dua kelas pada input space. Pola dari dua kelas dapat berupa +1 dan -1 dan bermacam alternatif garis pemisah. Margin merupakan jarak antar hyperplane dan pola terdekat dari setiap kelas, dimana pola yang terdekat disebut sebagai *support vector*[21].

6. K-Nearest Neighbor

Algoritma K-Nearest Neighbor (K-NN) merupakan suatu metode supervised learning dikarenakan hasil dari query instance perlu adanya klasifikasi dengan berdasarkan kartegori K-NN. Algoritma K-NN menentukan tetangga terdekat melalui target data dengan simbol K, apabila data training yang berdasarkan perhitungan jarak yang kurang akurat maka mencoba menentukan jarak yang digunakan memberikan hasil perhitungan yang terbaik[22]. Rumus jarak *Euclidean* sering digunakan guna mengukur jarak

antar data pada K-NN[23]. Rumus jarak *Euclidean* dapat ditunjukkan pada persamaan (1).

$$d(x_i, y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan :

d = Jarak

x_i = Sampel data

y_i = Data uji

i = Variabel

n = Dimensi data

7. Naïve Bayes

Naïve bayes adalah algoritma yang mampu mengklasifikasi sebuah variabel menggunakan sebuah metode statistik dan probabilitas. Metode probabilitas berguna dalam mencari peluang kemungkinan pada klasifikasi dengan melihat frekuensi pada data training setiap klasifikasi[24]. Probabilitas dihitung dengan menggunakan persamaan teorema bayes yang ditunjukkan pada persamaan (2).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2)$$

Keterangan :

$P(A|B)$ = Probabilitas A yang diberikan oleh B

$P(B|A)$ = Probabilitas B yang diberikan oleh A

$P(A)$ = Probabilitas kejadian A

$P(B)$ = Probabilitas kejadian B

8. Evaluasi

Perhitungan akurasi dari hasil yang telah didapatkan dari perhitungan model. Proses ini merupakan proses terakhir yang merepresentasikan kinerja yang telah digunakan guna menilai keberhasilan algoritma yang telah dipakai dalam mengklasifikasi. Hasil evaluasi dapat dilihat melalui *confussion matrix* yang memiliki empat kombinasi yaitu True Positive (TP), False Positive (FP), False Negative (FN), True Negative (TN)

Confussion matrix menghasilkan, akurasi untuk menghitung keakuratan model dalam mengklasifikasi pada persamaan (3). Presisi menghitung keakuratan model meprediksi benar positif dari keseluruhan yang diprediksi positif pada persamaan (4). Recall untuk menghitung keberhasilan model dalam menemukan prediksi benar positif dari seluruh data yang benar positif pada persamaan (5). F1-score menghitung rata-rata dari pesisi dan recall pada persamaan (6).

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

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$f1 - score = \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

Keterangan :

True Positive (TP) = Jumlah data yang bernilai positif dan diprediksi benar sebagai positif.

False Positive (FP) = Jumlah data yang bernilai negatif tetapi diprediksi sebagai positif.

False Negative (FN) = Jumlah data yang bernilai positif tetapi diprediksi sebagai negatif.

True Negative (TN) = Jumlah data yang bernilai negatif dan diprediksi benar sebagai negatif.

C. Hasil dan Pembahasan

1. Dataset

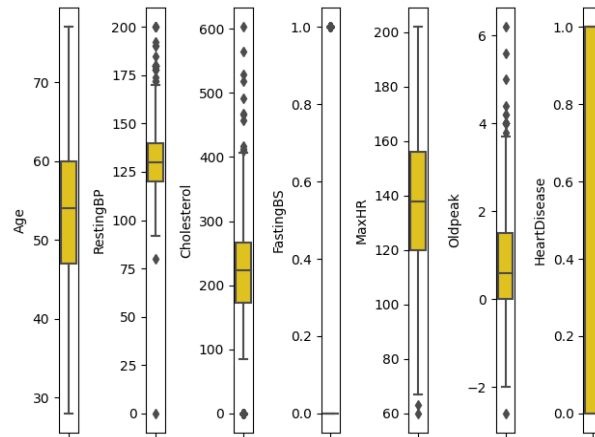
Tahap pertama yaitu pengumpulan data dimana data penyakit gagal jantung mempunyai 12 atribut yaitu :

Tabel 1. Keterangan Dataset

No	Atribut	Deskripsi
1	Age	Umur pasien
2	Sex	Jenis kelamin pasien, M: Male, F: Female
3	Chest Pain Type	Tipe nyeri dada, ATA: Atypical Angina, Non Angina Pain, ASY: Asymptomatic
4	Resting BP	Tekanan darah [mm Hg]
5	Cholesterol	Kolesterol serum [mm/dl]
6	FastingBS	Gula darah puasa [1: jika FastingBs > 120 mg/dl, 0: jika tidak]
7	Resting ECG	Hasil dari elektrokardiogram [Normal: normal, ST: memiliki kelainan gelombang ST-T (inversi gelombang T atau depresi ST > 0,05 mV), LVH: menunjukkan kemungkinan atau pasti hipertrofi ventrikel kiri menurut kriteria Estes]
8	MaxHR	Nilai maksimum detak jantung
9	Exercise Angina	Angina yang disebabkan olahraga [Y: ya, N: tidak]
10	Oldpeak	Oldpeak = ST [nilai ukur depresi]
11	ST_Slope	Kemiringan segmen ST [Up: upsloping, Flat: flat, Down: downsloping]
12	Heart Disease	Hasil penyakit gagal jantung [0: normal, 1: penyakit jantung]

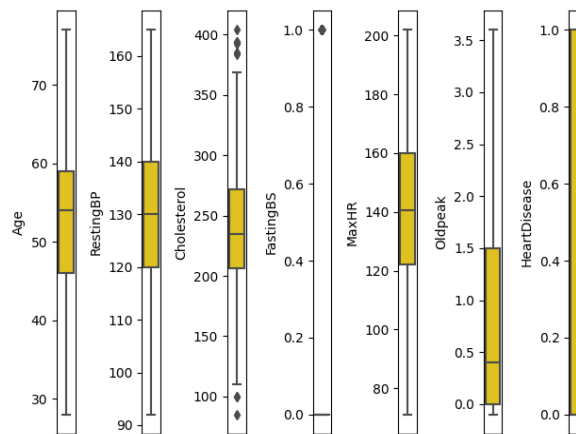
2. Preprocessing

Data yang digunakan pada penelitian ini tidak ada ada kosong dan duplikasi data setelah dilakukan pengecekan. Setelah pengecekan data kosong dan duplikasi data, dilanjut pada pengecekan outlier pada data.



Gambar 2. Outlier data

Pada Gambar 2 menunjukkan beberapa fitur pada data terdapat outlier setelah dilakukan pengecekan dengan *Interquartile Range* (IQR). Outlier dapat dilihat dari nilai yang disimbolkan dengan titik-titik dimana nilai tersebut berada jauh diluar jangkauan sesuai dengan *box plot* pada Gambar 2. Fitur yang terdapat outlier adalah fitur RestingBP, Cholesterol, MaxHR, dan Oldpeak. Dengan begitu maka dilakukan penghapusan outlier. Setelah dilakukan penghapusan outlier dapat dilihat pada *box plot* Gambar 3.



Gambar 3. Penghapusan Outlier

Setelah dilakukan penghapusan outlier data yang semula berjumlah 918 data menjadi 690 data. Setelah penghapusan outlier dilakukan lagi pengecekan data kosong, namun dalam pengecekan tidak ditemukan adanya data kosong. Selanjutnya tahap encoding dengan menggunakan *Label Encoder* pada fitur Sex, ChestPainType, Resting ECG, ExerciseAngina, dan ST_Slope.

3. Forward Selection

Seleksi fitur menggunakan metode *forward selection* dengan nilai tingkat signifikan 0.05 atau kepercayaan 95%. Seleksi *forward selection* menghasilkan fitur yang dianggap paling berpengaruh di antara semua fitur. Pada data yang digunakan fitur-fitur yang terseleksi dengan *forward selection* yaitu Age, Sex, ChestPainType, ExerciseAngina, Oldpeak, ST_Slope seperti yang ditunjukkan

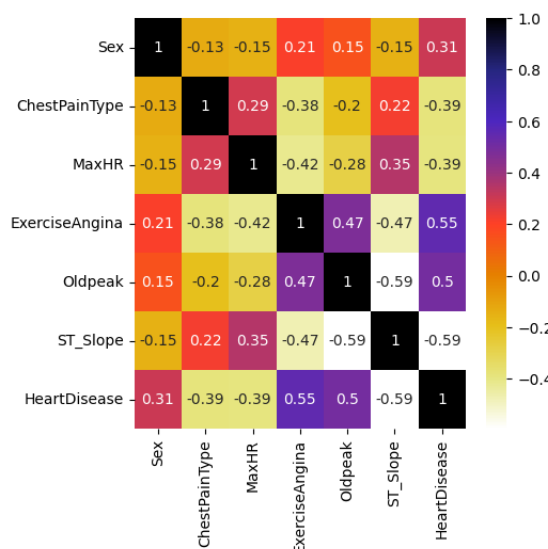
pada Gambar 4. Fitur yang terseleksi akan menjadi variabel independen dan untuk variabel dependen adalah fitur HeartDisease.

`['ST_Slope', 'ExerciseAngina', 'Sex', 'ChestPainType', 'Age', 'Oldpeak']`

Gambar 4. Seleksi Fitur Forward Selection

4. Pearson Corellation

Seleksi fitur menggunakan metode *Pearson corellation* dengan seleksi berdasarkan nilai korelasi varibel dengan variabel target. Fitur yang menjadi variabel target adalah HeartDisease. Nilai korelasi yang digunakan terhadap target yaitu lebih dari 0.3. Dengan begitu, fitur yang terseleksi dengan *Pearson corellation* terdiri dari Sex, ChestPainType, MaxHR, ExerciseAngina, Oldpeak, ST_Slope. Fitur yang terseleksi akan menjadi variabel independen dan untuk variabel dependen adalah fitur HeartDisease. Fitur yang terseleksi dapat dilihat hubungan antar variabelnya dengan grafik *heatmap* seperti yang ditunjukkan pada Gambar 5.



Gambar 5. Heatmap Selesi Fitur Pearson Corellation

5. Evaluasi

Pada penelitian ini menggunakan google colab dengan bahasa python. Data split yang digunakan guna membagi data training dan data testing dengan perbandingan 80:20 yang menghasilkan *accuracy*, *precision*, *recall*, dan *f1-score*. Adapun nilai parameter yang digunakan pada penelitian ini dengan algoritma SVM, KNN, dan Naïve Bayes ditunjukkan pada Tabel 2.

Tabel 2. Parameter Model

Algoritma	Parameter
SVM	Kernel = RBF
KNN	K = 7
Naïve Bayes	No parameter

Adapun hasil yang diperoleh dari komparasi algoritma SVM, KNN, dan Naïve Bayes tanpa menggunakan *feature selection forward selection* dan *Pearson corellation* ditunjukkan pada Tabel 3. Pada komparasi algoritma SVM, KNN, dan Naïve Bayes tanpa *feature selection* menggunakan semua fitur menjadi variabel independen dan fitur HeartDisease sebagai variabel dependen pada penelitian ini.

Tabel 3. Hasil Klasifikasi Algoritma Tanpa Feature Selection

Algoritma	Accuracy	Precision	Recall	F1-Score
SVM	88.41 %	91.04 %	85.92 %	88.41 %
KNN	87.68 %	90.91 %	84.51 %	87.59 %
Naïve Bayes	86.23 %	90.62 %	81.69 %	85.93 %

Berdasarkan hasil yang diperoleh dari klasifikasi menggunakan algoritma SVM, KNN, dan Naïve Bayes tanpa *feature selection* dapat dilihat perbandingannya. *Accuracy* tertinggi tanpa menggunakan *feature selection* dengan nilai 88.41%. Nilai *accuracy* tertinggi tanpa menggunakan *feature selection* dihasilkan algoritma SVM. Nilai *accuracy* terendah yang dihasilkan tanpa menggunakan *feature selection* 86.23%. Hasil *accuracy* terendah dihasilkan algoritma Naïve Bayes. Selain itu, nilai *precision* tertinggi yang dihasilkan 91.04%. Hasil *precision* tertinggi tanpa menggunakan *feature selection* dihasilkan algoritma SVM. Nilai *precision* terendah dihasilkan algoritma Naïve Bayes 86.23%. *Recall* tertinggi yang dihasilkan 85.92% pada algoritma SVM, sedangkan *recall* terendah dihasilkan algoritma Naïve Bayes dengan nilai 81.69%. *F1-score* tertinggi tanpa menggunakan *feature selection* yaitu 88.41% pada algoritma SVM dan *f1-score* terendah dihasilkan algoritma Naïve Bayes yaitu 85.93%.

Dengan demikian, komparasi algoritma SVM, KNN, dan Naïve Bayes tanpa menggunakan *feature selection*, algoritma yang paling unggul dibandingkan algoritma lainnya adalah algoritma SVM. Selain itu, penelitian ini menerapkan *feature selection forward selection* dan *Pearson corellation*. Hasil dari penerapan *feature selection* dapat ditunjukkan pada Tabel 4.

Tabel 4. Hasil Klasifikasi Algoritma Menggunakan Feature Selection

Algoritma	Feature Selection	Accuracy	Precision	Recall	F1-Score
SVM	Forward Selection	89.86 %	92.54 %	87.32 %	89.96 %
	Pearson Corellation	89.86 %	92.54 %	87.32 %	89.86 %
KNN	Forward Selection	88.41 %	92.31 %	84.51 %	88.24 %
	Pearson Corellation	90.58 %	93.94 %	87.32 %	90.51 %
Naïve Bayes	Forward Selection	87.68 %	90.91 %	84.51 %	87.59 %
	Pearson Corellation	86.96 %	92.06 %	81.69 %	86.57 %

Berdasarkan hasil klasifikasi algoritma SVM, KNN, dan Naïve Bayes dengan *feature selection forward selection* dan *Pearson corellation* dapat dilihat perbandingannya. Pada *feature selection forward selection* dengan algoritma SVM, KNN, dan Naïve Bayes *accuracy* tertinggi yaitu pada algoritma SVM dengan nilai 89.86%. *Precision* dengan *forward selection* algoritma SVM menghasilkan nilai *precision* tertinggi 92.54%. Hasil *precision* terendah dengan nilai 90.91% dihasilkan pada algoritma Naïve Bayes. *Recall* tertinggi menggunakan *forward*

selection 87.32% pada algoritma SVM, sedangkan pada algoritma KNN dan Naïve Bayes menghasilkan *recall* yang sama 84.51%. hasil *f1-score* menggunakan *forward selection* tertinggi 89.96% pada algoritma SVM dan terendah 87.59% pada algoritma Naïve Bayes.

Adapun pada penerapan *feature selection Pearson corellation* pada algoritma SVM, KNN, dan Naïve Bayes *accuracy* tertinggi 90.58%. *Accuracy* tertinggi pada algoritma KNN dan terendah pada algoritma Naïve Bayes dengan nilai 86.96%. *Precision* tertinggi pada algoritma KNN 93.94% dan terendah 92.06% pada algoritma Naïve Bayes. *Recall* tertinggi dengan nilai 87.32% pada algoritma SVM dan KNN. Nilai *recall* terendah 81.69% pada algoritma Naïve Bayes. Hasil *f1-score* tertinggi yaitu pada algoritma KNN dengan nilai 90.51%, sedangkan *f1-score* terendah dengan nilai 86.57% pada algoritma Naïve Bayes.

Dengan demikian berdasarkan hasil klasifikasi dengan penerapan *forward selection* dan *Pearson corellation* sama-sama dapat meningkatkan hasil *accuracy*. Pada penerapan *forward selection* algoritma SVM dapat meningkatkan *accuracy* 1,45%, algoritma KNN dapat meningkat 0,73%, dan algoritma Naïve Bayes meningkat 1,45% dibandingkan tanpa menggunakan *feature selection*. Penggunaan *Pearson corellation* meningkatkan *accuracy* pada algoritma SVM 1,45%, pada algoritma KNN meningkat 2,90%, dan pada algoritma Naïve Bayes meningkat 0,73% dibandingkan tanpa menggunakan *feature selection*. Algoritma yang paling unggul dibandingkan dengan algoritma lainnya pada *forward selection* adalah algoritma SVM, sedangkan pada *Pearson corellation* algoritma yang paling unggul adalah algoritma KNN. Dengan begitu, komparasi *feature selection* dari *forward selection* dan *Pearson corellation*, *feature selection* yang paling baik berdasarkan peningkatan hasil *accuracy* adalah *feature selection Pearson corellation*.

D. Simpulan

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan, pada komparasi algoritma SVM, KNN, Naïve Bayes tanpa menggunakan *feature selection* dan menggunakan *feature selection*. Algoritma SVM menjadi algoritma yang paling unggul dari algoritma lainnya tanpa menggunakan *feature selection* menghasilkan nilai *accuracy* 88.41%, *precision* 91.04%, *recall* 85.92%, dan *f1-score* 88.41%.

Pada penerapan *feature selection forward selection* dan *Pearson corellation* sama-sama dapat meningkatkan hasil akurasi baik pada algoritma SVM, KNN, dan Naïve Bayes. Algoritma SVM menjadi algoritma paling unggul pada penerapan *forward selection* dengan nilai *accuracy* 89.86%, *precision* 92.54%, *recall* 87.32%, *f1-score* 89.96%. Pada penerapan *Pearson corellation* algoritma KNN menjadi algoritma paling unggul dengan nilai *accuracy* 90.58%, *precision* 93.94%, *recall* 87.32%, dan *f1-score* 90.51%. Penerapan *feature selection* berdasarkan peningkatan *accuracy* dibandingkan tanpa menggunakan *feature selection Pearson corellation* menjadi *feature selection* terbaik pada penelitian ini.

E. Ucapan Terima Kasih

Terima kasih kepada Universitas Amikom Yogyakarta telah mendukung penuh penelitian ini.

F. Referensi

- [1] D. Karlia Destiani and D. Qori Utama, "KLASIFIKASI SINYAL ECGAGAL JANTUNG MENGGUNAKAN WAVELET DAN JST PROPAGASI BALIK DENGAN MODIFIKASI GRADIEN KONJUGAT POLAK-RIBIERE HEART FAILURE ECG SIGNAL CLASSIFICATION USING WAVELET AND ANN BACKPROPAGATION WITH POLAK-RIBIERE CONJUGATE GRADIENT MODIFICATION METHOD."
- [2] V. B. Aydemir *et al.*, "Classification of Decompensated Heart Failure from Clinical and Home Ballistocardiography," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 5, pp. 1303–1313, 2020, doi: 10.1109/TBME.2019.2935619.
- [3] R. Arisandi, "Perbandingan Model Klasifikasi Random Forest Dengan Resampling Dan Tanpa Resampling Pada Pasien Penderita Gagal Jantung," *J. Gaussian*, vol. 12, no. 1, pp. 136–145, 2023, doi: 10.14710/j.gauss.12.1.136-145.
- [4] F. Febby, A. Arjuna, and M. Maryana, "Dukungan Keluarga Berhubungan dengan Kualitas Hidup Pasien Gagal Jantung," *J. Penelit. Perawat Prof.*, vol. 5, no. 2, pp. 691–702, 2023, doi: 10.37287/jppp.v5i2.1537.
- [5] N. I. S. Lilik and I. Budiono, "Risiko Kematian Pasien Gagal Jantung Kongestif (GJK): Studi Kohort Retrospektif Berbasis Rumah Sakit," *Indones. J. Public Heal. Nutr.*, vol. 1, no. 3, pp. 388–395, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/IJPHN>
- [6] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, pp. 1–16, 2020, doi: 10.1186/s12911-020-1023-5.
- [7] A. P. Ayudhitama and U. Pujiyanto, "Analisa 4 Algoritma Dalam Klasifikasi Liver Menggunakan Rapidminer," *J. Inform. Polinema*, vol. 6, no. 2, pp. 1–9, 2020, doi: 10.33795/jip.v6i2.274.
- [8] Y. Pratama, A. Prayitno, D. Nazrian, N. Aini, Y. Rizki, and E. Rasywir, "Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor," *Bull. Comput. Sci. Res.*, vol. 3, no. 1, pp. 52–56, 2022, doi: 10.47065/bulletincsr.v3i1.203.
- [9] Prabiantissa Citra Nurina, "1818-5505-1-Pb," *Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform.*, vol. 1, no. 1, pp. 263–268, 2021.
- [10] Salamet Nur Himawan, Rendi, and Nur Budi Nugraha, "Feature Selection Menggunakan Algoritma Meta-Heuristik," *J. Pract. Comput. Sci.*, vol. 2, no. 2 SE-, pp. 84–89, Mar. 2023, doi: 10.37366/jpcs.v2i2.2289.
- [11] A. Salam, F. B. Nugroho, and J. Zeniarja, "Implementasi Algoritma K-Nearest Neighbor Berbasis Forward Selection Untuk Prediksi Mahasiswa Non Aktif Universitas Dian Nuswantoro Semarang," *JOINS (Journal Inf. Syst.)*, vol. 5, no. 1, pp. 69–76, 2020, doi: 10.33633/joins.v5i1.3351.
- [12] N. T. Romadloni and Hilman F Pardede, "Seleksi Fitur Berbasis Pearson Correlation Untuk Optimasi Opinion Mining Review Pelanggan," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 505–510, 2019, doi: 10.29207/resti.v3i3.1189.
- [13] H. D. Hekmatyar, W. A. Saputra, and C. Ramdani, "Klasifikasi Pneumonia Dengan Deep Learning Faster Region Convolutional Neural Network Arsitektur VGG16 dan ResNet50," *InComTech J. Telekomun. dan Komput.*, vol.

- 12, no. 3, p. 186, Dec. 2022, doi: 10.22441/incomtech.v12i3.15112.
- [14] I. N. Abrar and A. Abdullah, "Klasifikasi Penyakit Liver Menggunakan Metode Elbow Untuk Menentukan K Optimal pada Algoritma K-Nearest Neighbor (K-NN)," vol. 12, pp. 218–228, 2023.
- [15] N. P. Della Tirta Yanti and I. M. D. P. Asana, "Sistem Klasifikasi Pengajuan Kredit dengan Metode Support Vector Machine (SVM)," *J. Sist. Cerdas*, vol. 6, no. 2 SE-Articles, pp. 123–133, Aug. 2023, [Online]. Available: <https://apic.id/jurnal/index.php/jsc/article/view/287>
- [16] F. Rezkika, B. N. Sari, and A. Susilo, "Klasifikasi Masa Tunggu Alumni Untuk Mendapatkan Pekerjaan Menggunakan Algoritma C4 . 5," *Progresif J. Ilm. Komput.*, vol. Vol. 17, pp. 95–106, 2021.
- [17] H. Harafani and H. A. Al-Kautsar, "Meningkatkan Kinerja K-Nn Untuk Klasifikasi Kanker Payudara Dengan Forward Selection," *J. Pendidik. Teknol. dan Kejuru.*, vol. 18, no. 1, p. 99, 2021, doi: 10.23887/jptk-undiksha.v18i1.29905.
- [18] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
- [19] A. F. Aryuni, A. G. Putrada, and M. Abdurrohman, "Klasifikasi Penumpang Naik Dan Turun Dengan Sensor Load Cell Menggunakan Ekstraksi Fitur Dan Support Vector Machine," *e-Proceeding Eng.*, vol. 8, no. 2, pp. 3197–3208, 2021.
- [20] R. Yohannes and M. E. Al Rivan, "Klasifikasi Jenis Kanker Kulit Menggunakan CNN-SVM," *J. Algoritma.*, vol. 2, no. 2, pp. 133–144, 2022, doi: 10.35957/algoritme.v2i2.2363.
- [21] Ainani Shabrina Febrianti, Tri Arief Sardjono, and Atar Fuady Babgei, "Klasifikasi Tumor Otak pada Citra Magnetic Resonance Image dengan Menggunakan Metode Support Vector Machine," *J. Tek. ITS*, vol. Vol. 9, No, no. 1, 2020.
- [22] J. Homepage, S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," 2021.
- [23] V. Alvian, D. Hidayatullah, A. Nilogiri, H. Azizah, and A. Faruq, "Klasifikasi Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor (KNN) Pada SMA Negeri 2 Situbondo Classification Of Achieving Students Using K-Nearest Neighbor (KNN) Method At SMA Negeri 2 Situbondo," 2022. [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>
- [24] T. Imandasari, E. Irawan, A. Perdana Windarto, A. Wanto, and S. A. Tunas Bangsa Pematangsiantar Jln Jendral Sudirman Blok No, "Prosiding Seminar Nasional Riset Information Science (SENARIS) Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air," *Pros. Semin. Nas. Ris. Inf. Sci.*, no. September, pp. 750–761, 2019.