

Penerapan *Backward Elimination* Untuk Seleksi Fitur Pada Algoritma *K-Nearest Neighbour* Untuk Klasifikasi Penyakit Gagal Jantung

Indi Rosifatul Amilia^{1*}, Hardian Oktavianto², Ginanjar Abdurrahman³

¹²³Program Studi Teknik Informatika, Universitas Muhammadiyah Jember

Email: indirosi5@gmail.com^{1*}, hardian@unmuhjember.ac.id², abdurrahmanginanjar@unmuhjember.ac.id³

ABSTRAK

Gagal jantung adalah penyakit kardiovaskular yang disebabkan oleh jantung yang memompa darah ke seluruh tubuh dan mengganggu fungsi sistem sirkulasi fisiologis. Data di Indonesia tahun 2018 menunjukkan bahwa gagal jantung termasuk dalam 10 penyakit tidak menular di Indonesia dan mendapatkan hasil 229.696 (0,13%) orang menderita gagal jantung. Penyakit gagal jantung tidak dapat disembuhkan, tetapi banyak kasus dapat dicegah dan sebagian besar pasien dapat diobati secara efektif untuk meningkatkan kualitas hidup dan kelangsungan hidup. Seiring dengan perkembangan zaman dalam dunia kesehatan terjadi juga perkembangan yang cukup pesat. Salah satunya adalah pemanfaatan *Machine learning* dan *Data mining* dalam dunia kesehatan. Metode klasifikasi pada penelitian ini adalah membandingkan antara algoritma *K-Nearest Neighbour* tanpa seleksi fitur dan *K-Nearest Neighbour* dengan menggunakan seleksi fitur *Backward Elimination*. Hasil yang didapatkan dari penelitian ini adalah pada algoritma *K-Nearest Neighbour* tanpa seleksi fitur didapatkan hasil akurasi sebesar 94,56%, presisi 93,87% dan *Recall* 95,55%. Pada algoritma *K-Nearest Neighbour* dengan menggunakan seleksi fitur *Backward Elimination* mendapatkan hasil akurasi sebesar 98,33%, presisi 97,94% dan *recall* 98,63%.

Kata Kunci: Klasifikasi, Gagal Jantung, *K-Nearest Neighbour*, *Backward Elimination*

ABSTRACT

Heart failure is a cardiovascular disease caused by the heart pumping blood throughout the body and disrupting the function of the physiological circulatory system. Data in Indonesia in 2018 showed that heart failure was included in the 10 non-communicable diseases in Indonesia and resulted in 229,696 (0.13%) people suffering from heart failure. Heart failure is incurable, but many cases can be prevented and most patients can be treated effectively to improve quality of life and survival. Along with the times in the world of health there is also a fairly rapid development. One of them is the use of Machine learning and Data mining in the world of health. The classification method in this study is to compare the K-Nearest Neighbour algorithm without feature selection and K-Nearest Neighbour using the Backward Elimination feature selection. The results obtained from this study are the K-Nearest Neighbour algorithm without feature selection, the results obtained are 94.56% accuracy, 93.87% precision and 95.55% Recall. In the K-Nearest Neighbour algorithm using the Backward Elimination feature selection, the results are 98.33% accuracy, 97.94% precision and 98.63% Recall.

Keywords: Classification, Heart Failure, *K-Nearest Neighbour*, *Backward Elimination*

1. PENDAHULUAN

Gagal jantung adalah penyakit kardiovaskular yang disebabkan oleh jantung yang memompa darah ke seluruh tubuh dan mengganggu fungsi sistem sirkulasi fisiologis. Prevalensi gagal jantung di Asia Tenggara mencapai 3 kali lipat jika dibandingkan dengan negara Eropa dan Amerika yaitu sebesar 4.5–6.7% : 0.5–2% (Lam, 2015). Angka kematian akibat gagal jantung tinggi yaitu mencapai 50% dalam 5 tahun. Penting untuk mengetahui tentang faktor-faktor yang mempengaruhi kematian akibat penyakit ini agar dapat dilakukan pencegahan dini. Salah satu faktor yang dapat meningkatkan kejadian gagal jantung adalah gaya hidup yang tidak sehat dan kemampuan perawatan diri. Penyakit gagal jantung tidak dapat disembuhkan, tetapi banyak kasus dapat dicegah dan sebagian besar pasien dapat diobati secara efektif untuk meningkatkan kualitas hidup dan kelangsungan hidup. Seiring dengan perkembangan zaman dalam dunia kesehatan terjadi juga perkembangan yang cukup pesat. Salah satunya adalah pemanfaatan *machine learning* dan *data mining* dalam dunia kesehatan.

Klasifikasi merupakan bagian dari *data mining*, dimana *data mining* merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database* (Sumarlin, 2015). Klasifikasi adalah teknik pengolahan data yang membagi suatu data objek menjadi beberapa kelas sesuai dengan jumlah kelas yang dibutuhkan. Salah satu algoritma metode klasifikasi yaitu *K-Nearest*

Neighbour. Algoritma *K-Nearest Neighbour* adalah metode untuk melakukan klasifikasi pada objek berdasarkan data latih yang jaraknya paling dekat dengan objek. Keakuratan algoritma *K-Nearest Neighbour* dipengaruhi oleh atribut yang tidak relevan atau jika bobot kriteria tersebut tidak setara dengan relevansinya terhadap klasifikasi. Algoritma *K-Nearest Neighbour* juga memiliki keunggulan pelatihan yang sangat cepat dan sederhana sehingga mudah dipelajari. Namun ada kekurangan pada algoritma *K-Nearest Neighbour* yaitu perlu penentuan nilai *K* dan untuk pemilihan atribut terbaik. Untuk pemilihan atribut terbaik maka dibutuhkan optimasi yaitu dengan menggunakan seleksi fitur.

Backward Elimination merupakan salah satu tahap yang dapat dilakukan dalam melakukan seleksi atribut pada tahap preprocessing. *Backward Elimination* yang berfungsi untuk mengoptimalkan kinerja suatu model dengan sistem kinerja mundur yang digunakan untuk memilih atribut paling relevan. Metode *Backward Elimination* memberikan kinerja yang lebih baik jika dibandingkan dengan metode statistik signifikan dalam tahap seleksi. Kinerja terbaik dibuktikan dengan sensitivitas, spesifisitas, dan akurasi yang tinggi. Pada penelitian ini *Backward Elimination* digunakan untuk seleksi fitur yang bertujuan untuk meningkatkan nilai akurasi.

Pada penelitian terdahulu yang berjudul Implementasi *Data mining* untuk Deteksi Penyakit Ginjal Kronis (PGK) Menggunakan *K-Nearest Neighbour* dengan *Backward Elimination* menggunakan atribut yang telah diseleksi menggunakan *Backward Elimination* berhasil menekan biaya pemeriksaan hingga 73,36%, dan untuk hasil pendeteksi penyakit menggunakan Algoritma *K-Nearest Neighbour* menghasilkan nilai akurasi sebesar 99,25% (Gamadarenda & Waspada, 2018). Pada penelitian yang berjudul Klasifikasi Penyakit Jantung Menggunakan Metode *K-Nearest Neighbour* menggunakan data penyakit jantung dan mendapatkan *K* terbaik pada *K=6* dengan akurasi sebesar 85% (Hasran, 2020). Melihat dari penelitian-penelitian sebelumnya perbedaan dengan penelitian sebelumnya yaitu dapat dilihat dari data atribut yaitu 13 atribut yang digunakan dan pada tujuan penelitian.

Berdasarkan penjelasan diatas, maka penulis tertarik untuk mengklasifikasi penyakit gagal jantung dengan judul “Penerapan *Backward Elimination* Untuk Seleksi Fitur Pada Algoritma *K-Nearest Neighbour* Untuk Klasifikasi Penyakit Gagal Jantung”.

2. KAJIAN PUSTAKA

A. Gagal Jantung

Gagal jantung merupakan salah satu penyakit kardiovaskuler yang paling sering terjadi di seluruh dunia yang mengakibatkan tingginya angka mortalitas, morbiditas dan juga berdampak secara finansial terutama bagi lanjut usia. Rehospitalisasi merupakan masalah umum yang sering terjadi pada pasien gagal jantung yang sebagian besar disebabkan oleh keterlambatan dalam pengenalan gejala, pengobatan dan ketidakpatuhan diet serta kurangnya pengetahuan dan keterampilan dalam melakukan perawatan diri (Prihatiningsih & Sudyasih, 2018).

B. *Data mining*

Data mining merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. *Data mining* mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data. *Data mining* juga dikenal sebagai knowledge discovery, knowledge extraction, data/pattern analysis, information harvesting. *Data mining* bertujuan untuk menemukan pola yang sebelumnya tidak diketahui. Jika pola-pola tersebut telah diperoleh maka dapat digunakan untuk menyelesaikan berbagai macam permasalahan (Arhami & Nasir, 2020).

Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar.

C. Klasifikasi

Klasifikasi adalah proses penemuan model yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan supaya bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han et al., 2006). Model yang diturunkan didasarkan pada analisis dari training data.

Klasifikasi adalah contoh *task machine learning* tipe *supervised* untuk melakukan prediksi. Dalam klasifikasi kita dapat menentukan orang atau objek kedalam suatu kategori tertentu. Informasi tentang objek sebelumnya digunakan sebagai bahan algoritma untuk mendapatkan rule.

D. K-Nearest Neighbour

K-Nearest Neighbour adalah pendekatan untuk menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan penyesuaian bobot dari banyak fitur yang ada (Simpson, 2019). Tujuan dari algoritma *K-Nearest Neighbour* yaitu untuk klasifikasi data berdasarkan variabel serta sampel data training. Klasifikasi tersebut menggunakan voting yang paling banyak antara klasifikasi objek k yang sudah ditentukan. Dalam menentukan pencarian jarak terdekat akan digunakan perhitungan *Euclidean Distance*. Untuk perhitungan rumus jarak *Euclidean Distance* dari *K-Nearest Neighbour* dapat dijelaskan pada persamaan (Kusrini, 2018) adalah sebagai berikut:

$$d_i = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

Keterangan:

p : Dimensi data

i : Atribut data

x : Data sampel / data latih

y : Data uji / data testing

d : Jarak

E. Backward Elimination

Backward Elimination merupakan metode yang dapat menghilangkan atribut yang tidak signifikan dari model. Berfungsi sebagai seleksi atribut dimana memanfaatkan regresi statistik untuk mengetahui kedekatan setiap kombinasi atribut dengan target. Semakin kecil significance level, maka semakin ketat pemilihan atribut yang akan terpilih sehingga semakin sedikit atribut yang terpilih sebagai model.

Metode Backward merupakan metode regresi yang baik karena pada metode ini menggambarkan variabel perilaku respon yang terbaik dengan memilih variabel penjelas dari banyak variabel penjelas yang tersedia pada data. Membentuk persamaan regresi linier dengan metode backward (Noya et al., 2019)

- Meregresikan seluruh variabel bebas dengan variabel terikat, selanjutnya lihat pada uji parsial dan bandingkan nilai *p-value* setiap variabel dengan nilai taraf signifikan.
- Keluarkan variabel yang memiliki nilai *p-value* < taraf signifikan dari persamaan regresi.
- Meregresikan variabel bebas yang tersisa dengan variabel terikat.
- Perhatikan hasil uji parsialnya. Jika terdapat nilai *p-value* dari variabel bebas yang < taraf signifikan maka kembali ke langkah b. Sedangkan jika nilai *p-value* setiap variabel bebas > taraf signifikan maka lanjutkan ke langkah berikutnya.
- Menentukan model regresi yang signifikan secara parsial dan serentak.
- Melakukan pengujian asumsi klasik untuk regresi linier berganda.
- Menentukan nilai koefisien determinasi untuk model dengan metode *backward*

F. Confusion Matrix

Confusion Matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep *data mining* (Wisdayani et al., 2019). *Confusion Matrix* berisi informasi yang aktual tentang perkiraan klasifikasi yang dihasilkan sistem dengan hasilnya menggunakan data ke dalam bentuk sebuah matrix. Berikut ini tabel *Confusion Matrix* (Salam et al., 2020). Untuk menghitung nilai akurasi, presisi, *Recall* menggunakan persamaan sebagai berikut:

a. Akurasi

Akurasi adalah perbandingan kasus yang benar dibagi dengan jumlah seluruh kasus. Berikut ini rumus untuk mencari nilai akurasi:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

b. Presisi

Presisi adalah nilai untuk menunjukkan tingkat keberhasilan sebuah model dalam mengenali suatu objek. Berikut ini rumus untuk mencari nilai presisi:

$$Presisi = \frac{TP}{TP + FP}$$

c. Recall

Recall adalah nilai untuk menunjukkan tingkat keberhasilan sebuah model dalam mengenali yang bukan suatu objek. Berikut ini rumus untuk mencari nilai presisi:

$$Recall = \frac{TP}{TP + FN}$$

3. METODE PENELITIAN

A. Rancangan Penelitian

Rancangan penelitian yang digunakan dalam penelitian ini antara lain mengumpulkan data, seleksi fitur, klasifikasi.



Gambar 1. Alur Penelitian

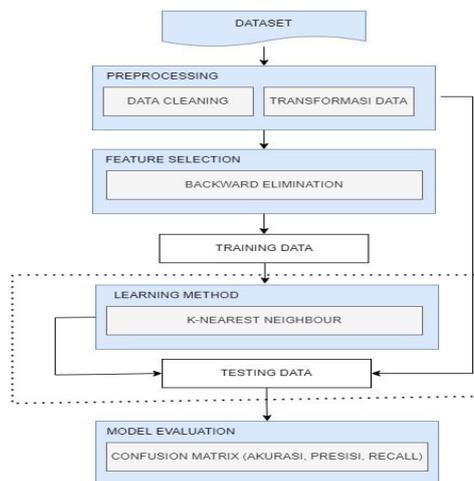
B. Pengumpulan Data

Pada penelitian ini, data yang digunakan adalah data penyakit gagal jantung. Data tersebut diambil melalui situs *UCI Machine learning*, yang menyediakan berbagai macam data untuk keperluan *machine learning*. Data diambil secara manual dengan mengunduh melalui halaman *website* pada *UCI machine learning*. Data yang digunakan pada penelitian ini merupakan data hasil klasifikasi penyakit

gagal jantung pada tahun 2020. Data penyakit gagal jantung tersebut dikategorikan menjadi dua kelas output yaitu meninggal dan selamat. Pada data tersebut terdapat 12 atribut yaitu *age*, *anemia*, *creatinine phosphokinase*, *diabetes*, *ejection fraction*, *high blood pressure*, *platelets*, *serum creatinine*, *serum sodium*, *sex*, *smoking*, *time*. Data yang digunakan berjumlah sebanyak 299 data.

C. Pengolahan Awal Data

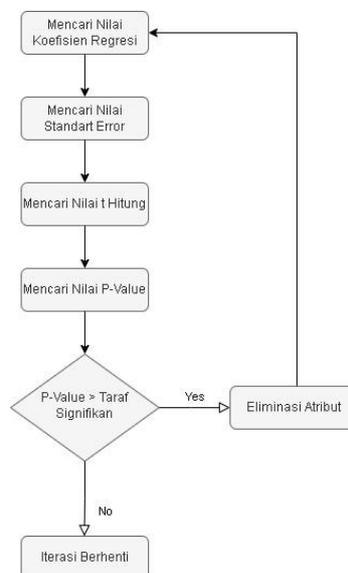
Pada proses ini dilakukan untuk membersihkan data dari *noise*. Pada tahapan pengolahan awal data diawali dengan *cleaning* data yaitu membersihkan data dari data-data yang hilang (*missing value*) atau data yang tidak lengkap agar hasil yang didapatkan dari hasil klasifikasi bisa menjadi lebih akurat. Selanjutnya dilakukan transformasi data dengan menyamakan format pada setiap atribut dari dataset dengan mengubah kategori menjadi inisial angka. Contohnya pada kelas *output Death Event* mengubah *value* pasien meninggal dengan inisial 1 dan pasien selamat menjadi 0.



Gambar 2. Alur Pengolahan Data

D. Metode

Metode yang diusulkan adalah menggunakan teknik *Backward Elimination* untuk menyeleksi fitur atau atribut dari variabel bebas terhadap variabel terikatnya yang kemudian dikombinasikan dengan menggunakan algoritma *K-Nearest Neighbour*. Hasil yang didapatkan adalah hasil kinerja algoritma berupa akurasi, presisi, dan *recall*. Kemudian hasil yang didapatkan akan dibandingkan antara algoritma *K-Nearest Neighbour* dengan menggunakan seleksi fitur *Backward Elimination* dan algoritma *K-Nearest Neighbour* tanpa menggunakan seleksi fitur *Backward Elimination*.



Gambar 3. Alur *Backward Elimination*

Berikut adalah alur perhitungan Teknik *Backward Elimination*:

1. Mencari nilai Koefisien Regresi
2. Menghitung nilai Standart Error pada masing-masing atribut
3. Mencari nilai t-hitung untuk masing-masing atribut
4. Selanjutnya menentukan nilai *p-value* berdasarkan pada nilai t-tabel
5. Menentukan nilai *p-value* terbesar dari masing-masing atribut
6. Selanjutnya membandingkan nilai antara *p-value* terbesar dengan taraf signifikan, apabila nilai *p-value* lebih besar dari taraf signifikannya maka atribut dengan *p-value* terbesar tersebut akan dieliminasi dan tahap selanjutnya adalah mengulangi langkah pertama kembali sampai nilai *p-value* pada setiap atribut yang tersisa kurang dari taraf signifikannya.

E. Eksperimen dan Pengujian Metode

Eksperimen yang dilakukan membandingkan antara hasil kinerja algoritma *K-Nearest Neighbour* dengan menggunakan seleksi fitur *Backward Elimination* dan tanpa menggunakan seleksi fitur *Backward Elimination*. Teknik *Backward Elimination* akan menyeleksi atribut yang kurang berpengaruh terhadap output dengan mengeliminasi atribut yang kurang berpengaruh tersebut. Untuk eksperimen pada data penelitian menggunakan aplikasi Rapid Miner untuk mengolah data.

F. Evaluasi dan Validasi

Pengukuran akurasi, presisi, dan *Recall* dilakukan dengan menggunakan confusion matrix. Penggunaan data training yang digunakan berjumlah 299 data yang diambil secara acak pada data asli. Selanjutnya dilakukan proses perhitungan *Backward Elimination* untuk seleksi fiturnya. Proses validasi dilakukan pada kumpulan dataset penyakit gagal jantung.

Proses terakhir pada penelitian ini adalah melakukan evaluasi terhadap hasil klasifikasi yang telah dilakukan. Proses evaluasi ini dilakukan untuk mendapatkan nilai akurasi dari penggunaan algoritma *K-Nearest Neighbour* yang menggunakan seleksi fitur *Backward Elimination* dan yang tanpa menggunakan seleksi fitur. Dari hasil nilai yang didapatkan akan dipilih nilai akurasi yang terbaik antara menggunakan seleksi fitur *Backward Elimination* dan tanpa menggunakan seleksi fitur.

4. HASIL DAN PEMBAHASAN

Data yang akan digunakan berjumlah 299 data penyakit gagal jantung. Dari 299 data tersebut selanjutnya akan dilakukan normalisasi data agar data bersih dari *noise*. Setelah dilakukan normalisasi data selanjutnya akan dilakukan dua skenario yaitu dengan menggunakan seleksi fitur *Backward Elimination* dan tanpa menggunakan seleksi fitur. Setelah dilakukan seleksi fitur maka tahap selanjutnya adalah tahap klasifikasi menggunakan algoritma *K-Nearest Neighbour*.

Berikut hasil yang didapatkan dari hasil klasifikasi algoritma *K-Nearest Neighbour* $k = 3, 5, 7$ tanpa menggunakan *Backward Elimination*.

Tabel 1 Hasil Klasifikasi KNN tanpa *Backward Elimination*

<i>K-KNN</i>	Akurasi	Presisi	<i>Recall</i>
3	94,56%	93,87%	95,55%
5	70,71%	70,30%	65,88%
7	70,71%	65,88%	70,30%

Berikut hasil yang didapatkan dari hasil klasifikasi algoritma *K-Nearest Neighbour* $k = 3, 5, 7$ dengan menggunakan *Backward Elimination*

Tabel 2 Hasil Klasifikasi KNN dengan *Backward Elimination*

<i>K-KNN</i>	Akurasi	Presisi	<i>Recall</i>
3	98,33%	97,94%	98,63%
5	87,03%	88,03%	89,7%
7	84,94%	86,08%	82,25%

Berdasarkan hasil klasifikasi algoritma *K-Nearest Neighbour* tanpa menggunakan seleksi fitur *Backward Elimination* pada $k=3$ mendapatkan hasil akurasi sebesar 94,56%, presisi sebesar 93,87, dan *recall* sebesar 95,55%. Pada $k=5$ didapatkan hasil akurasi sebesar 70,71%, presisi sebesar 70,30%, dan *recall* sebesar 65,88%. Pada $k=7$ didapatkan hasil akurasi sebesar 70,71%, presisi sebesar 65,88%, dan *recall* sebesar 70,30%. Sedangkan hasil klasifikasi algoritma *K-Nearest Neighbour* dengan menggunakan seleksi fitur *Backward Elimination* pada $k=3$ mendapatkan hasil akurasi sebesar 98,33%, presisi sebesar 97,94, dan *recall* sebesar 98,63%. Pada $k=5$ didapatkan hasil akurasi sebesar 87,03%, presisi sebesar 88,03%, dan *recall* sebesar 89,70%. Pada $k=7$ didapatkan hasil akurasi sebesar 84,94%, presisi sebesar 86,08%, dan *recall* sebesar 82,25%.

5. KESIMPULAN

A. Kesimpulan

1. Didapatkan hasil akurasi tertinggi pada klasifikasi penyakit gagal jantung dengan menggunakan algoritma *K-Nearest Neighbour* tanpa seleksi fitur *Backward Elimination* yaitu sebesar 94,56% pada $k = 3$, sedangkan nilai akurasi tertinggi dengan menggunakan seleksi fitur *Backward Elimination* yaitu sebesar 98,33% pada $k = 3$.
2. Didapatkan hasil presisi tertinggi pada klasifikasi penyakit gagal jantung dengan menggunakan algoritma *K-Nearest Neighbour* tanpa seleksi fitur *Backward Elimination* yaitu sebesar 93,87% pada $k = 3$, sedangkan nilai presisi tertinggi dengan menggunakan seleksi fitur *Backward Elimination* yaitu sebesar 97,94% pada $k = 3$.
3. Didapatkan hasil *Recall* tertinggi pada klasifikasi penyakit gagal jantung dengan menggunakan algoritma *K-Nearest Neighbour* tanpa seleksi fitur *Backward Elimination* yaitu sebesar 95,55% pada $k = 3$, sedangkan nilai *recall* tertinggi dengan menggunakan seleksi fitur *Backward Elimination* yaitu sebesar 97,94% pada $k = 3$.
4. Didapatkan hasil akurasi, presisi, dan *recall* tertinggi yaitu dengan menggunakan seleksi fitur *Backward Elimination* dengan akurasi sebesar 98,33%, *recall* sebesar 97,94%, dan *recall* sebesar 98,63%.

B. Saran

Berdasarkan penelitian yang telah dijabarkan di atas, adapun beberapa saran yang diberikan untuk penelitian selanjutnya adalah sebagai berikut:

1. Diharapkan untuk peneliti selanjutnya agar menggunakan data penyakit gagal jantung terbaru untuk mendapatkan hasil yang lebih baik.
2. Diharapkan untuk peneliti selanjutnya bisa menerapkan pembuatan sistem klasifikasi penyakit gagal jantung dengan penelitian ini sebagai acuan.
3. Diharapkan untuk peneliti selanjutnya menggunakan data yang lebih banyak, karena semakin banyak data yang digunakan semakin baik mesin mempelajari model yang diolah dan dapat memungkinkan untuk mendapatkan hasil yang lebih optimal.

6. DAFTAR PUSTAKA

- Arhami, M., & Nasir, M. (2020). *Data Mining-Algoritma dan Implementasi*. ANDI.
- Gamadarenda, I. W., & Waspada, I. (2018). Implementasi Data Mining untuk Deteksi Penyakit Ginjal Kronis (PGK) menggunakan K-Nearest Neighbor (KNN) dengan Backward Elimination. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 7(2).
- Han, J., Pei, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2nd ed.). Morgan Kaufmann.

- Hasran. (2020). Klasifikasi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor. *Indonesian Journal of Data and Science*, 1(1).
- Lam, C. S. P. (2015). Heart failure in Southeast Asia: facts and numbers. In *ESC Heart Failure* (Vol. 2, Issue 2, pp. 46–49). Wiley-Blackwell. <https://doi.org/10.1002/ehf2.12036>.
- Noya, M. S., Delsen, V., Patty, H. W. M., Lalurmele, N. L., & Matematika, J. (2019). Model of Multiple Linear Regression with Backward and Forward Method (Case Study: Local Tax Revenue of Ambon City in 2007-2016). *Journal of Statistics and Its Applications*, 1(1).
- Prihatiningsih, D., & Sudyasih, T. (2018). Perawatan Diri Pada Pasien Gagal Jantung. *JURNAL PENDIDIKAN KEPERAWATAN INDONESIA*, 4(2). <https://doi.org/10.17509/jpki.v4i2.13443>.
- Salam, A., Nugroho, F. B., & Zeniarja, J. (2020). Implementasi Algoritma K-Nearest Neighbor Berbasis Forward Selection Untuk Prediksi Mahasiswa Non Aktif Universitas Dian Nuswantoro Semarang. *JOINS (Journal of Information System)*, 5(1), 69–76. <https://doi.org/10.33633/joins.v5i1.3351>.
- Simpem, M. S. M. I. W. (2019). Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba. *Seminar Sistem Informasi Dan Teknologi Informasi*, 8(1), 1–10.
- Sumarlin, S. (2015). Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM Sumarlin STIKOM Uyelindo Kupang. *JSINBIS (Jurnal Sistem Informasi Bisnis)*, 5(1), 52–62.
- Wisdayani, D. S., Manfaati Nur, I., & Wasono, R. (2019). *Perbandingan Algoritma K-Nearest Neighbor Dan Naive Bayes Untuk Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Di Kabupaten Pati Jawa Tengah*. <http://repository.unimus.ac.id>.