

Prediksi Mortalitas Gagal Jantung Menggunakan PCA dan K-Nearest Neighbors: Analisis Komparatif Metrik Jarak

Feralia Fitri, Navessa Julieth

Politeknik Negeri Batam, Batam, Indonesia

INFORMASI ARTIKEL	ABSTRAK
<p>Sejarah Artikel: Diterima: Revisi: Diterima: Dipublikasi:</p> <p>Kata Kunci: Heart Failure, Prediction, Mortality, KNN, PCA</p> <p>*Penulis Korespondensi: navessajulieth@gmail.com</p>	<p>Gagal jantung merupakan kondisi ketika jantung tidak mampu memompa darah secara optimal untuk memenuhi kebutuhan metabolik tubuh. Prediksi yang akurat diperlukan untuk mendukung intervensi medis yang tepat waktu. Penelitian ini mengkaji penggunaan metode K-Nearest Neighbors (KNN) untuk mengklasifikasikan luaran pasien gagal jantung berdasarkan data tetangga terdekat pada data pelatihan. Metode ini dikombinasikan dengan Principal Component Analysis (PCA) guna memprediksi mortalitas pasien. Hasil kajian menegaskan bahwa KNN merupakan metode yang sederhana dan efektif dalam analisis data medis.</p> <p>ABSTRACT <i>Heart failure is a condition in which the heart is unable to pump blood optimally to meet the body's metabolic demands. Accurate prediction is essential to support timely medical intervention. This study examines the use of the K-Nearest Neighbors (KNN) method to classify heart failure patient outcomes based on nearest-neighbor data from the training set. The method is combined with Principal Component Analysis (PCA) to predict patient mortality, demonstrating that KNN is a simple and effective approach for medical data analysis.</i></p>

PENDAHULUAN

Gagal jantung merupakan kondisi ketika jantung tidak mampu memompa darah secara efisien untuk memenuhi kebutuhan metabolik tubuh. Pada kondisi ini, meskipun darah tetap kembali ke jantung, kemampuan jantung untuk memompa darah ke seluruh jaringan tubuh menurun sehingga distribusi oksigen dan nutrisi menjadi tidak optimal. Akibatnya, organ dan jaringan tubuh tidak dapat menjalankan fungsinya secara normal.

Prediksi risiko mortalitas pada pasien gagal jantung secara akurat memerlukan penerapan berbagai pemeriksaan diagnostik yang tepat. Namun, keterbatasan keahlian tenaga medis serta kompleksitas data klinis dapat menyebabkan ketidakakuratan dalam proses penilaian risiko. Salah satu pendekatan yang dapat digunakan untuk membantu proses prediksi adalah pemanfaatan data rekam medis. Rekam medis mampu merekam riwayat kondisi pasien serta indikator klinis yang berguna dalam analisis biostatistik dan pengambilan keputusan medis. Dalam bidang komputasi, metode *machine learning* dapat dimanfaatkan untuk mengolah data tersebut guna memprediksi luaran klinis serta mengidentifikasi faktor-faktor penting yang perlu diperhatikan dalam rekam medis pasien.

K-Nearest Neighbors (KNN) merupakan salah satu metode *machine learning* dengan pendekatan *supervised learning* yang relatif mudah dipahami dan diimplementasikan. Metode ini termasuk dalam kategori *lazy learning*, di mana data pelatihan disimpan terlebih dahulu dan proses klasifikasi dilakukan ketika data uji diberikan. Klasifikasi pada KNN ditentukan berdasarkan kedekatan jarak antara data uji dan data pelatihan, dengan hasil prediksi ditentukan oleh kategori mayoritas dari sejumlah tetangga terdekat [4][5].

Dalam analisis data medis, reduksi dimensi sering kali diperlukan untuk mengatasi kompleksitas data dan mengurangi redundansi fitur. Principal Component Analysis (PCA) merupakan salah satu teknik yang efektif untuk tujuan tersebut karena mampu mentransformasikan data ke dalam sejumlah komponen utama yang merepresentasikan sebagian besar variasi data. Pada penelitian ini, metode PCA digunakan untuk mereduksi dimensi data, sedangkan KNN digunakan sebagai metode klasifikasi untuk memprediksi mortalitas pada pasien gagal jantung.

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma K-Nearest Neighbors (KNN) dalam memprediksi mortalitas pada pasien gagal jantung. Hasil penelitian menunjukkan bahwa model yang diusulkan berhasil mencapai tingkat akurasi sebesar 88%, yang mengindikasikan potensi metode ini untuk diterapkan sebagai alat bantu dalam analisis data klinis dan pengambilan keputusan medis.

METODE PENELITIAN

A. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) merupakan metode klasifikasi yang menentukan kelas suatu data berdasarkan mayoritas kelas dari sejumlah tetangga terdekat yang diukur menggunakan metrik jarak. Pada penelitian ini digunakan tiga metrik jarak, yaitu Euclidean, Manhattan, dan Chebyshev. Jarak Euclidean mengukur jarak geometris antar titik dalam ruang multidimensi, jarak Manhattan menghitung jumlah selisih absolut tiap dimensi, sedangkan jarak Chebyshev menggunakan selisih maksimum pada salah satu dimensi. Ketiga metrik tersebut digunakan untuk mengevaluasi pengaruh perbedaan ukuran jarak terhadap kinerja model [6].

B. Data Normalization

Normalisasi fitur dilakukan sebagai tahap prapemrosesan untuk mengatasi perbedaan skala antar fitur yang dapat memengaruhi perhitungan jarak pada algoritma KNN. Pada penelitian ini, normalisasi dilakukan menggunakan metode *StandardScaler*, yang mentransformasikan data sehingga memiliki nilai rata-rata 0 dan simpangan baku 1 [7].

C. Feature Extraction with PCA

Principal Component Analysis (PCA) digunakan sebagai teknik reduksi dimensi data. Pada penelitian ini, jumlah fitur direduksi menjadi dua komponen utama dengan menetapkan parameter $n_components = 2$, dengan tujuan untuk menurunkan kompleksitas model serta meminimalkan risiko *overfitting*, khususnya pada data dengan jumlah fitur yang besar. Selain itu, hasil reduksi dimensi digunakan untuk memvisualisasikan data dalam bentuk grafik dua dimensi guna mengamati pola dan sebaran data berdasarkan dua komponen utama tersebut.

D. Evaluating Model Performance

Penelitian ini mengevaluasi kinerja model K-Nearest Neighbors (KNN) dalam memprediksi mortalitas pasien gagal jantung. Evaluasi kinerja model dilakukan dengan meninjau

beberapa metrik utama yang diperoleh dari *classification report* dan *confusion matrix*. Berikut penjelasan dari metrik-metrik tersebut:

1. **Precision** merupakan proporsi prediksi positif yang benar (*true positive*) terhadap seluruh prediksi positif yang dihasilkan oleh model. Metrik ini mengukur sejauh mana kasus yang diprediksi sebagai positif benar-benar relevan atau sesuai dengan kondisi aktual.

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)}$$

2. **Recall** merupakan proporsi prediksi positif yang benar (*true positive*) terhadap seluruh kasus positif yang sebenarnya. Metrik ini mengukur kemampuan model dalam mengidentifikasi dan menangkap seluruh kasus positif secara tepat.

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)}$$

3. **F1-Score** merupakan nilai rata-rata harmonik dari *precision* dan *recall* yang memberikan keseimbangan antara kedua metrik tersebut. Metrik ini digunakan ketika *precision* dan *recall* memiliki tingkat kepentingan yang sama dalam mengevaluasi kinerja model.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

4. **Accuracy** merupakan proporsi prediksi yang benar terhadap keseluruhan jumlah prediksi yang dihasilkan oleh model. Nilai akurasi diperoleh dengan membagi jumlah prediksi yang tepat, baik *true positive* maupun *true negative*, dengan total seluruh prediksi.

$$Akurasi = \frac{True\ Positives\ (TP) + True\ Negatives\ (TN)}{Totas\ kasus\ (TP + TN + FP + FN)}$$

5. **Confusion Matrix** merupakan sebuah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan nilai prediksi terhadap nilai aktual. *Confusion matrix* terdiri dari empat komponen utama, yaitu:

- **TP (True Positive):** jumlah kasus positif yang diprediksi dengan benar sebagai positif.
- **TN (True Negative):** jumlah kasus negatif yang diprediksi dengan benar sebagai negatif.
- **FP (False Positive):** jumlah kasus negatif yang keliru diprediksi sebagai positif.
- **FN (False Negative):** jumlah kasus positif yang keliru diprediksi sebagai negatif.

Dengan menggunakan metrik-metrik tersebut, penelitian ini dapat mengevaluasi tingkat efektivitas model KNN dalam memprediksi mortalitas pasien gagal jantung serta melakukan perbaikan model apabila diperlukan.

HASIL DAN PEMBAHASAN

Hasil evaluasi model K-Nearest Neighbors (KNN) menunjukkan kinerja yang baik dalam memprediksi mortalitas pada pasien gagal jantung, dengan tingkat akurasi sebesar 0,88. Nilai *precision*, *recall*, dan *F1-score* yang berada pada kisaran 0,88 mengindikasikan bahwa model memiliki kemampuan klasifikasi yang konsisten serta andal dalam mengidentifikasi risiko mortalitas pada pasien.

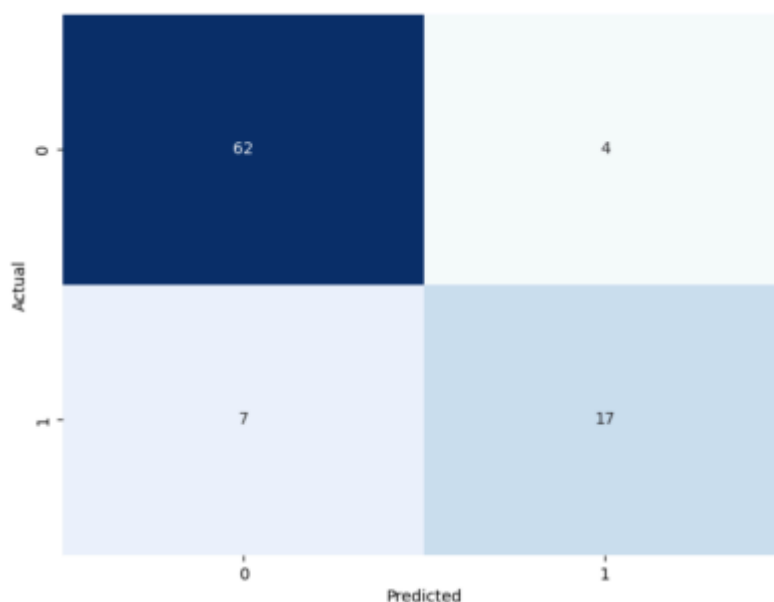
A. Hasil

Hasil evaluasi model K-Nearest Neighbors (KNN) menunjukkan bahwa model mampu memprediksi mortalitas pada pasien gagal jantung dengan tingkat akurasi sebesar 0,88 pada parameter $K = 20$ dan metrik jarak Chebyshev. Nilai precision yang diperoleh masing-masing sebesar 0,90 untuk kelas 0 (tidak meninggal) dan 0,81 untuk kelas 1 (meninggal), sedangkan nilai recall sebesar 0,94 untuk kelas 0 dan 0,71 untuk kelas 1.

Nilai F1-score sebesar 0,92 pada kelas 0 dan 0,76 pada kelas 1 menunjukkan keseimbangan yang baik antara precision dan recall. Berdasarkan confusion matrix, diperoleh 62 true negative, 4 false positive, 7 false negative, dan 17 true positive. Hasil ini menunjukkan bahwa model KNN dengan parameter tersebut mampu menghasilkan prediksi yang akurat dengan tingkat kesalahan yang relatif rendah.

1. Confusion Matrix

Hasil ini merupakan hasil terbaik dari eksperimen dalam menentukan nilai parameter K dan metrik jarak yang paling optimal untuk model KNN.



Gambar 1. Confusion Matrix pada Model KNN dengan Nilai K dan Metrik Jarak Terbaik

Accuracy	Precision	Recall	F1-score
88%	87%	88%	88%

Tabel 1. Hasil Optimal: $K = 20$, Metrik Chebyshev

Data	Precision	Recall	F1-score	Support
0	90%	94%	92%	66
1	81%	71%	76%	24

Tabel 2. Laporan Klasifikasi Hasil Optimal ($K=20$, Metrik Chebyshev)

Hasil terbaik dari eksperimen menggunakan model K-Nearest Neighbors (KNN) dengan parameter $K = 20$ dan metrik jarak Chebyshev menunjukkan tingkat akurasi sebesar 0,88. Model ini juga menghasilkan nilai F1-score, precision, dan recall masing-masing sebesar 0,88, 0,87, dan 0,88. Berdasarkan confusion matrix, terdapat 62 prediksi benar pada kelas 0 (tidak meninggal) dan 17 prediksi benar pada kelas 1 (meninggal), dengan 4 prediksi salah pada kelas 0 dan 7 prediksi salah pada kelas 1.

Evaluasi lebih lanjut menunjukkan bahwa model memiliki kinerja yang lebih baik dalam mengidentifikasi kelas 0, dengan nilai *precision* sebesar 0,90 dan *recall* sebesar 0,94. Sementara itu, untuk kelas 1 diperoleh nilai *precision* sebesar 0,81 dan *recall* sebesar 0,71. Secara keseluruhan, model ini menunjukkan kinerja yang stabil dan andal dalam memprediksi mortalitas pada pasien gagal jantung.

2. Evaluasi Kinerja Model

K	Metric Distance											
	euclidean				manhattan				chebyshev			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
1	82%	82%	82%	82%	81%	81%	81%	81%	83%	84%	83%	83%
2	83%	83%	83%	83%	84%	84%	84%	84%	82%	81%	82%	81%
3	83%	83%	83%	83%	83%	84%	83%	83%	82%	82%	82%	82%
4	83%	83%	83%	83%	84%	84%	84%	84%	83%	83%	83%	83%
5	82%	82%	82%	82%	82%	82%	82%	82%	82%	82%	82%	82%
6	87%	86%	87%	86%	86%	85%	86%	85%	86%	85%	86%	85%
7	83%	84%	83%	83%	82%	82%	82%	82%	84%	84%	84%	84%
8	86%	85%	86%	85%	83%	83%	83%	83%	84%	84%	84%	84%
9	84%	86%	84%	85%	83%	86%	83%	84%	82%	83%	82%	83%
10	84%	84%	84%	84%	83%	83%	83%	83%	83%	83%	83%	83%
11	80%	81%	80%	80%	86%	86%	86%	86%	80%	81%	80%	80%
12	82%	83%	82%	82%	86%	86%	86%	86%	80%	81%	80%	80%
13	80%	81%	80%	80%	84%	86%	84%	85%	79%	81%	79%	79%
14	86%	85%	86%	85%	86%	86%	86%	86%	81%	81%	81%	81%
15	83%	85%	83%	84%	83%	86%	83%	84%	81%	82%	81%	81%
16	83%	84%	83%	84%	87%	87%	87%	87%	84%	84%	84%	84%
17	81%	83%	81%	82%	84%	85%	84%	85%	82%	83%	82%	83%
18	86%	85%	86%	85%	84%	84%	84%	84%	84%	84%	84%	84%
19	82%	83%	82%	83%	82%	83%	82%	82%	86%	85%	86%	85%
20	86%	85%	86%	85%	83%	83%	83%	83%	88%	87%	88%	88%

Tabel 3. Nilai Akurasi K=1 hingga 20 dengan Metrik Euclidean, Manhattan, dan Chebyshev

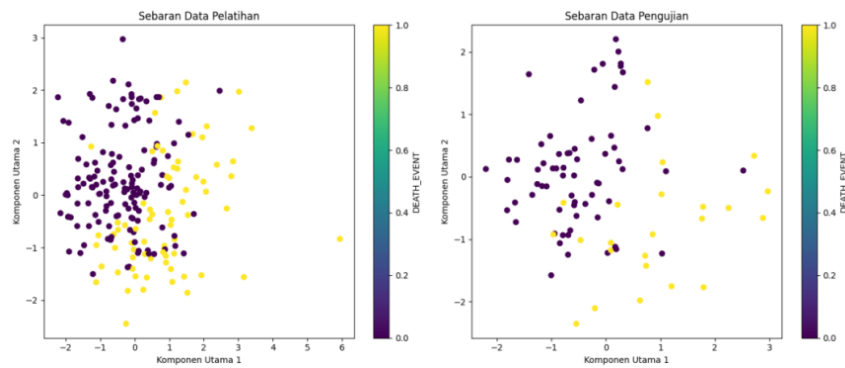
K	Metric Distance											
	euclidean				manhattan				chebyshev			
	TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN
1	58	8	8	16	57	9	8	16	58	8	7	17
2	61	5	10	14	61	5	9	15	62	4	12	12
3	59	7	8	16	58	8	7	17	58	8	8	16
4	61	5	10	14	61	5	9	15	61	5	10	14
5	58	8	8	16	58	8	8	16	59	7	9	15
6	63	3	9	15	62	4	9	15	62	4	9	15
7	58	8	7	17	58	8	8	16	60	6	8	16
8	62	4	9	15	60	6	9	15	60	6	8	16
9	56	10	4	20	54	12	3	21	56	10	6	18
10	59	7	7	17	59	7	8	16	59	7	8	16
11	55	11	7	17	58	8	5	19	55	11	7	17
12	57	9	7	17	59	7	6	18	56	10	8	16
13	55	11	7	17	56	10	4	20	54	12	7	17
14	60	6	7	17	58	8	5	19	58	8	9	15
15	56	10	5	19	55	11	4	20	56	10	7	17
16	57	9	6	18	60	6	6	18	59	7	7	17
17	55	11	6	18	58	8	6	18	56	10	6	18
18	61	5	8	16	60	6	8	16	60	6	8	16
19	56	10	6	18	57	9	7	17	60	6	7	17
20	61	5	8	16	59	7	8	16	62	4	7	17

Tabel 4. Confusion Matrix untuk K=1 hingga 20 dengan Metrik Euclidean, Manhattan, dan Chebyshev

- a. Model KNN dengan Parameter K = 1 dan Metrik Jarak Euclidean
 Dengan menggunakan parameter K = 1 dan metrik jarak Euclidean, model *K-Nearest Neighbors* (KNN) mencapai nilai akurasi, *precision*, *recall*, dan *F1-score* sebesar 82%, yang menunjukkan kemampuan yang cukup baik dalam memprediksi kelas pada dataset. Namun, masih terdapat beberapa kasus *false positive* dan *false negative*. Kondisi ini menunjukkan bahwa kinerja model masih dapat ditingkatkan melalui eksplorasi nilai K yang berbeda atau dengan menggunakan metrik jarak lain guna meningkatkan performa prediksi.
- b. Model KNN dengan Parameter K = 1 dan Metrik Jarak Manhattan
 Dengan menggunakan parameter K = 1 dan metrik jarak Manhattan, model *K-Nearest Neighbors* (KNN) mencapai tingkat akurasi sebesar 81%, dengan nilai *precision*, *recall*, dan *F1-score* yang juga berada pada angka 81%. Hasil ini menunjukkan kinerja yang sedikit lebih rendah dibandingkan dengan penggunaan metrik jarak Euclidean. Meskipun model tetap efektif dalam memprediksi label kelas, masih terdapat beberapa kasus *false positive* dan *false negative*. Hal ini mengindikasikan bahwa kinerja model masih dapat ditingkatkan melalui penyesuaian nilai K atau dengan mempertimbangkan penggunaan metrik jarak alternatif guna meningkatkan akurasi dan keandalan prediksi.
- c. Model KNN dengan Parameter K = 1 dan Metrik Jarak Chebyshev
 Dengan menggunakan parameter K = 1 dan metrik jarak Chebyshev, model *K-Nearest Neighbors* (KNN) mencapai tingkat akurasi sebesar 83%, dengan nilai *precision*, *recall*, dan *F1-score* yang juga sebesar 83%. Hasil ini menunjukkan kinerja yang sebanding dengan metrik jarak Euclidean dan sedikit lebih baik dibandingkan metrik Manhattan pada dataset yang digunakan. Model ini menunjukkan kemampuan prediksi yang baik, namun seperti pada penggunaan metrik lainnya, masih terdapat kasus *false positive* dan *false negative*. Oleh karena itu, peningkatan kinerja model masih dimungkinkan melalui eksplorasi nilai K yang berbeda atau dengan mempertimbangkan penggunaan metrik jarak alternatif guna meningkatkan akurasi dan ketahanan prediksi.

Hasil evaluasi menunjukkan adanya variasi kinerja model *K-Nearest Neighbors* (KNN) pada berbagai kombinasi nilai K dan metrik jarak (Euclidean, Manhattan, dan Chebyshev), dengan tingkat akurasi berkisar antara 79% hingga 88%. Nilai K yang terlalu kecil, seperti K = 1, menyebabkan *overfitting* yang ditunjukkan oleh fluktuasi nilai *precision*, *recall*, dan *F1-score*, sedangkan nilai K yang lebih besar menghasilkan model yang lebih stabil. Perbedaan kinerja antar metrik jarak relatif kecil, namun metrik Chebyshev menunjukkan performa yang lebih unggul pada kondisi tertentu, seperti pada K = 20. Temuan ini menjadi acuan penting dalam pemilihan parameter optimal pada penerapan model KNN.

3. Distribusi Data



Gambar 1. Visualisasi *scatter plot* data pelatihan dan data pengujian penerapan *Principal Component Analysis* (PCA)

a. Distribusi Data Pelatihan (Grafik Kiri)

Grafik ini menunjukkan distribusi data pelatihan setelah dilakukan transformasi menggunakan *Principal Component Analysis* (PCA). Sumbu horizontal (Komponen Utama 1) dan sumbu vertikal (Komponen Utama 2) merepresentasikan dua komponen utama yang dihasilkan oleh PCA. Kedua komponen tersebut merupakan kombinasi linear dari fitur-fitur asli yang mampu menjelaskan sebagian besar variasi data. Warna titik pada grafik menunjukkan nilai *DEATH_EVENT*, di mana warna kuning merepresentasikan nilai 1 (terjadi kematian) dan warna ungu merepresentasikan nilai 0 (tidak terjadi kematian). Gradasi warna antara ungu dan kuning menunjukkan probabilitas atau intensitas terjadinya *DEATH_EVENT*.

b. Distribusi Data Pengujian (Grafik Kanan)

Grafik ini menunjukkan distribusi data pengujian setelah dilakukan transformasi menggunakan PCA yang sama seperti pada data pelatihan. Serupa dengan grafik sebelah kiri, sumbu horizontal (Komponen Utama 1) dan sumbu vertikal (Komponen Utama 2) merepresentasikan dua komponen utama yang dihasilkan oleh PCA. Warna titik pada grafik ini juga menunjukkan nilai *DEATH_EVENT* dengan interpretasi yang sama seperti pada grafik pelatihan.

Secara keseluruhan, kedua grafik ini digunakan untuk memvisualisasikan sebaran data pelatihan dan data pengujian pada ruang komponen utama yang dihasilkan oleh PCA, serta distribusi variabel *DEATH_EVENT* pada masing-masing dataset. Visualisasi ini membantu dalam mengidentifikasi pola dan kluster data yang mungkin tidak terlihat pada ruang fitur asli.

B. Pembahasan

Model KNN menunjukkan peningkatan akurasi dan stabilitas kinerja seiring peningkatan nilai K dari 1 hingga 20, dengan metrik Chebyshev mencapai akurasi tertinggi 88% pada K=20. Pada rentang K=1–5, akurasi berada di kisaran 81–84% disertai false positive (FP) dan false negative (FN) yang relatif tinggi; peningkatan K mengurangi FP dan FN secara signifikan, menghasilkan TP=62, TN=17, FP=4, dan FN=7 pada K=20 dengan Chebyshev. Hal ini menggambarkan dinamika bias-variance dalam KNN, di mana nilai K yang lebih besar meminimalkan varians melalui agregasi tetangga. Metrik jarak memengaruhi performa: Chebyshev efektif pada K tinggi karena menekankan jarak maksimum, Euclidean mencapai 87% pada K=6–10, dan Manhattan stabil pada K=11–15 hingga 86%. Normalisasi data dan PCA mendukung pemisahan kelas yang jelas pada scatter plot train/test, sehingga meningkatkan generalisasi model. Evaluasi menunjukkan

precision, recall, dan F1-score optimal pada kombinasi tersebut, mengonfirmasi efektivitas KNN untuk deteksi gagal jantung dengan implikasi pada sistem pemantauan klinis.

KESIMPULAN

Penelitian ini menunjukkan bahwa model K-Nearest Neighbors (KNN) efektif dalam memprediksi mortalitas pada pasien gagal jantung, dengan akurasi yang signifikan dan stabilitas kinerja yang meningkat seiring penambahan nilai K serta variasi metrik jarak. Penelitian menggunakan 299 data pasien, dengan hasil terbaik pada $K = 20$ menggunakan metrik Chebyshev yang mencapai akurasi 88%, ditandai dengan penurunan signifikan false positive ($FP = 4$) dan false negative ($FN = 7$), serta true positive ($TP = 62$) dan true negative ($TN = 17$). Optimalisasi model juga dipengaruhi oleh proses normalisasi data dan ekstraksi fitur menggunakan PCA yang membantu memperjelas pola penting dalam data.

Daftar Pustaka

- [1] R. Andini and Y. P. Astuti, "Model infeksi HIV dengan pengaruh percobaan," *J. Ilm. Mat.*, vol. 9, no. 2, pp. 437–446, 2021. [Online]. Available: <https://media.neliti.com/media/publications/249234-model-infeksihiv-dengan-pengaruh-percob-b7e3cd43.pdf>
- [2] 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.
- [3] R. Burle, S. Gaurkhede, L. Dewangan, and U. Pacharaney, "Prediction of heart disease using machine learning algorithms," in *2024 IEEE Int. Conf. Interdiscip. Approaches Technol. Manag. Soc. Innov. (IATMSI)*, 2024, pp. 1–5, doi: 10.1109/IATMSI60426.2024.10502796.
- [4] N. R. Rajalakshmi, J. Santhosh, J. Arun Pandian, and M. Alkhouli, "Prediction of chronic heart disease using machine learning," in *Smart Innov., Syst. Technol.*, vol. 334, pp. 177–185, 2023, doi: 10.1007/978-981-19-8497-6_18.
- [5] K. N. N. Untuk, "Implementasi algoritma klasifikasi K-Nearest Neighbor," *Jurnal*, vol. 6, no. 2, pp. 118–127, 2021.
- [6] S. García, J. Luengo, and F. Herrera, "Instance selection," in *Data Preprocessing in Data Mining*, vol. 72, Springer, 2015, doi: 10.1007/978-3-319-10247-4_8.