

Implementasi *Multilayer Perceptron* Untuk Memprediksi Harapan Hidup Pada Pasien Penyakit Kardiovaskular

Wilda Imama Sabilla¹, Candra Bella Vista², Dhebys Suryani Hormansyah³

^{1,2,3}Jurusan Teknologi Informasi, Politeknik Negeri Malang

Jl. Soekarno Hatta No.9, Jatimulyo, Kec. Lowokwaru, Kota Malang, (0341)404424

¹wildaimama@polinema.ac.id, ²bellavista@polinema.ac.id, ³dhebys.suryani@gmail.com

Abstract

Cardiovascular disease is one of the leading causes of death in the world. The risk of death is important to predict to determine treatment or behavior and lifestyle changes in cardiovascular patients. Medical record data of cardiovascular patients can be used as input in predicting life expectancy. This study offers the construction of a life expectancy prediction system for cardiovascular patients. Prediction using multilayer perceptron method by testing various scenarios. In addition, feature selection methods, namely correlation based filter (CBF), linear discriminant analysis (LDA), and principal component analysis (PCA) are applied to obtain relevant features to improve classification performance. Based on the experiments conducted, the average accuracy using CBF and LDA feature selection is 84% and 84.7%, respectively. In the best trial, CBF is able to produce accuracy, precision, recall, and f-measure with value of 91.7% 85% 89.5% and 87.2%. Based on these results, it can be concluded that this prediction system is able to provide fairly accurate results.

Keywords: *cardiovascular, feature selection, multilayer perceptron, prediction*

Abstrak

Penyakit kardiovaskular menjadi salah satu penyakit yang menyumbang kematian terbesar di dunia. Resiko kematian penting untuk diprediksi untuk menentukan pengobatan ataupun perubahan perilaku dan gaya hidup pada pasien kardiovaskular. Data rekam medis pasien kardiovaskular dapat dimanfaatkan sebagai masukan dalam memprediksi harapan hidup. Penelitian ini menawarkan dibangunnya sistem prediksi harapan hidup bagi pasien kardiovaskular. Prediksi menggunakan metode multilayer perceptron dengan uji coba berbagai skenario. Selain itu, metode seleksi fitur yaitu correlation based filter (CBF), linear discriminant analysis (LDA), dan principal component analysis (PCA) diterapkan guna memperoleh fitur relevan untuk menaikkan performa klasifikasi. Berdasarkan uji coba yang dilakukan, rata-rata akurasi menggunakan seleksi fitur CBF dan LDA sebesar 84% dan 84,7%. Pada uji coba terbaik, CBF mampu menghasilkan nilai akurasi, precision, recall, dan f-measure yaitu 91,7% 85% 89,5% dan 87,2%. Berdasarkan hasil tersebut dapat disimpulkan bahwa sistem prediksi ini mampu memberikan hasil yang cukup akurat.

Kata kunci: *kardiovaskular, multilayer perceptron, prediksi, seleksi fitur*

1. PENDAHULUAN

Penyakit kardiovaskular, yang biasa dikenal dengan penyakit jantung, adalah penyakit yang terjadi ketika jantung mengalami gangguan. Gangguan tersebut bermacam-macam antara lain gangguan pada pembuluh darah jantung, katup jantung, irama jantung, maupun bawaan lahir [1]. Penyakit kardiovaskular merupakan penyakit tidak menular namun memiliki resiko kematian tinggi. Penyakit kardiovaskular mengakibatkan kematian dini sekitar 4% di negara berpenghasilan tinggi dan sebesar 42% di negara



berpenghasilan rendah [2]. Penyakit kardiovaskular menjadi penyebab kematian nomor satu di seluruh dunia dan menjadi penyebab dari lebih dari 30% kematian global. Diperkirakan pada tahun 2030 kematian global karena kardiovaskular meningkat menjadi 22 juta jiwa jika tidak segera diatasi [3]. Di Indonesia, jumlah penderita penyakit kardiovaskular naik dari tahun ke tahun. Berdasarkan data Riset Kesehatan Dasar (Riskesdas) tahun 2018, setidaknya 15 dari 1000 orang, atau sekitar 2.784.064 individu di Indonesia menderita penyakit jantung baik gangguan jantung maupun pembuluh darah jantung [4].

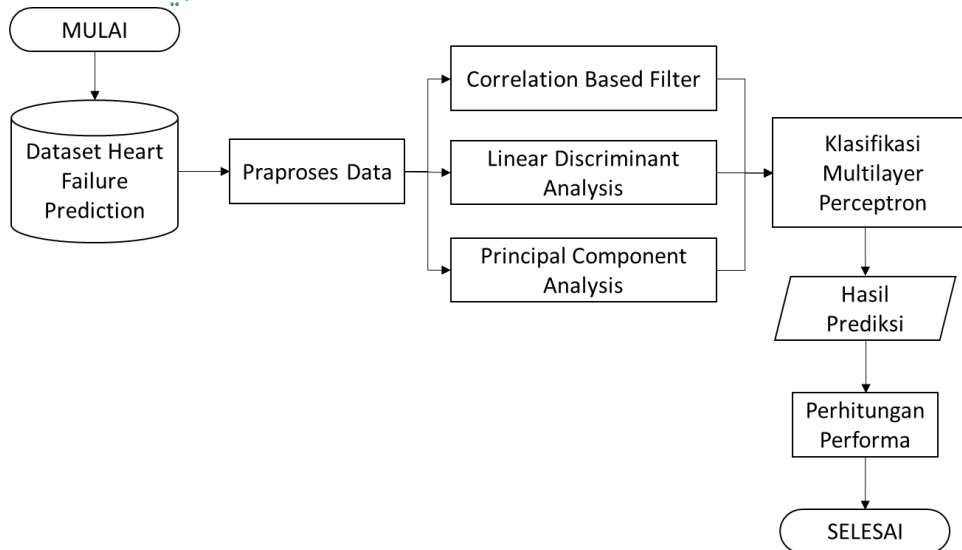
Tingginya angka kematian akibat penyakit kardiovaskular menjadi salah satu perhatian di bidang kesehatan secara global. Oleh karena itu penilaian resiko sangat penting. Penilaian resiko dapat digunakan untuk memprediksi resiko kematian atau harapan hidup. Prediksi resiko dapat digunakan untuk saran pengobatan ataupun perubahan gaya hidup dan perilaku penderita penyakit kardiovaskular [5]. Untuk mengatasi keterlambatan penanganan maupun diagnosis pasien, diperlukan suatu sistem yang dapat membantu dokter terlebih untuk dokter-dokter yang kurang berpengalaman [6]. Begitu juga dengan data rekam medis pasien penyakit kardiovaskular, data ini dapat dimanfaatkan untuk dianalisis ke dalam sistem yang berguna untuk memprediksi harapan hidup atau resiko kematian pada pasien kardiovaskular.

Terdapat beberapa penelitian mengenai sistem prediksi menggunakan data rekam medis pasien penyakit kardiovaskular yang menereapkan berbagai teknik pemodelan dan klasifikasi. Pada penelitian prediksi penyakit jantung koroner yang dilakukan oleh Oladimedji dkk, algoritma *naïve bayes* mendapatkan akurasi sebesar 86,67%. Sedangkan menggunakan algoritma C4.5, mereka memperoleh akurasi sebesar 83,70% [7]. Penelitian [8] mendapatkan akurasi 86,6% dalam sistem prediksi resiko kematian penyakit kardiovaskular dengan metode *associative classification*. Sedangkan penggunaan algoritma *random forest* dan *k-nearest neighbor* memperoleh akurasi masing-masing 77,75% dan 75,09% dalam system diagnosis penyakit jantung [9].

Penelitian ini akan mengimplementasikan metode *multilayer perceptron* untuk memprediksi resiko kematian atau harapan hidup pada pasien penyakit kardiovaskular. Prediksi menggunakan data rekam medis pasien yang akan diproses untuk mengetahui apakah pasien memiliki resiko tinggi atau rendah. Metode *multilayer perceptron* diharapkan dapat menghasilkan performa yang baik. Sebelum menerapkan *multilayer perceptron*, data akan diproses menggunakan tiga metode seleksi fitur yaitu *Correlation Based Filter (CBF)*, *Linear Discriminant Analysis (LDA)*, dan *Principal Component Analysis (PCA)*. Seleksi fitur diharapkan dapat memperbaiki tingkat ketepatan prediksi. Performa atau ketepatan prediksi akan dihitung berdasarkan nilai *akurasi*, *precision*, *recall*, dan *f-measure*.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan dengan membangun sistem prediksi harapan hidup pasien penyakit kardiovaskular berdasarkan alur yang dapat dilihat pada Gambar 1.



Gambar 1. Alur metode penelitian

Pada awalnya data yang ada akan dilakukan praproses untuk normalisasi data. Agar klasifikasi maksimal, dari data yang sudah dipraproses akan dilakukan seleksi fitur terlebih dahulu. Terdapat tiga metode yaitu LBF, LDA, dan PCA untuk memilih fitur terbaik. Data yang telah diseleksi fiturnya selanjutnya diklasifikasikan menggunakan *multilayer perceptron*. Data yang tersedia akan dibagi sebagai data latih dan data uji dengan rasio 80% dan 20%. Hasil klasifikasi berupa prediksi dari seorang pasien memiliki resiko kematian mendadak atau tidak berdasarkan rekam medis pada dataset. Terakhir hasil prediksi akan diukur melalui perhitungan performa dan dianalisis untuk mengetahui apakah sistem layak digunakan.

2.1. Data

Data yang digunakan pada penelitian ini adalah *dataset* “Heart Failure Prediction” yang merupakan *open access data* dari halaman web Kaggle (<https://www.kaggle.com/andrewmvd/heart-failure-clinical-data>). *Dataset* ini dikumpulkan oleh Davide Chicco dan Giuseppe Jurman berdasarkan rekam medis 299 pasien gagal jantung di Punjab, Pakistan [10]. Data ini dikumpulkan di Institut Kardiologi Faisalabad bekerja sama dengan Rumah Sakit Faisalabad selama April – Desember 2015.

Pasien terdiri dari 105 wanita dan 194 pria, dan usia mereka berkisar antara 40 dan 95 tahun. 299 pasien mengalami disfungsi sistolik ventrikel kiri dan memiliki gagal jantung sebelumnya yang menempatkan mereka di kelas III atau IV dari klasifikasi New York Heart Association (NYHA) dari

tahapan gagal jantung. Penjelasan dari fitur-fitur pada *dataset* dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi Data

Fitur	Keterangan	Satuan	Jangkauan nilai
Age	Umur pasien	Tahun	40-95
Anaemia	Berkurangnya sel darah merah (hemoglobin)	Boolean	0,1
High blood pressure	Apakah memiliki hipertensi?	Boolean	0,1
Creatinine Phosphokinase (CPK)	Level CPK di dalam darah	mcg/L	23-7861
Diabetes	Apakah pasien diabetes?	Boolean	0,1
Ejection fraction	Persentase darah yang meninggalkan jantung	Persen	14-80
Sex	Jenis kelamin pasien	Boolean	0,1
Platelets	Kandungan trombosit di dalam darah	Kiloplatelets/m L	25,01-850,00
Serum creatinine	Kandungan kreatinin di dalam darah	mg/dL	0,50-9,40
Serum sodium	Kandungan sodium di dalam darah	mEq/L	114-148
Smoking	Apakah pasien perokok?	Boolean	0,1
Time	Periode <i>follow-up</i>	Hari	4-285
Label/Kelas: Death Event	Apakah pasien meninggal selama periode <i>follow-up</i>	Boolean	0,1

2.2. Praproses

Nilai fitur yang pada *dataset* "Heart Failure Prediction" memiliki jangkauan nilai yang berbeda-beda. Untuk itu dilakukan normalisasi agar jangkauan data lebih seragam. Metode praproses yang digunakan adalah *standard scaler*. *Standard scaler* menghitung *mean* dan standar deviasi dari nilai fitur, mengurangi *mean* dari setiap nilai dan membagi hasilnya dengan standar deviasi [10].

2.3. Seleksi Fitur

Seleksi fitur adalah sebuah metode untuk memilih fitur dari sekumpulan fitur yang ada sehingga dimensi fitur dapat dikurangi. Pengurangan fitur dilakukan untuk memperoleh fitur yang relevan, hal ini dapat mempercepat proses klasifikasi serta meningkatkan akurasi dari algoritma klasifikasi [11]. Pada penelitian ini diterapkan tiga metode seleksi fitur yang berbeda yaitu CBF, LDA, dan PCA. Selanjutnya akan diperoleh tiga data yang berbeda dari masing-masing hasil seleksi fitur.

a) Correlation Based Feature

Correlation Based Feature (CBF) yaitu salah satu teknik pemilihan fitur menggunakan perhitungan statistik korelasi. CBF akan menghasilkan

correlation matrix yang menyimpan korelasi satu fitur terhadap fitur lainnya [12]. Nilai korelasi yang dihasilkan berada pada rentang [-1, 1]. Fitur yang memiliki nilai korelasi lebih tinggi terhadap kelas menjadi fitur yang akan dipilih. Parameter yang digunakan untuk menyeleksi fitur adalah nilai *threshold* yang dalam penelitian ini *threshold* yang digunakan adalah 0,1 (baik positif maupun negatif). Fitur-fitur yang diperoleh berdasarkan *threshold* yang digunakan yaitu 'age', 'ejection_fraction', 'serum_creatinine', 'serum_sodium', dan 'time'. Nilai korelasi fitur terhadap kelas 'death_event' berdasarkan *correlation matrix* ditunjukkan pada Tabel II.

Tabel 2. Nilai Korelasi Fitur terhadap Kelas

Fitur	Nilai Korelasi
Age	0,25
Anaemua	0,066
High blood preasure	0,079
Creatinine Phosphokinase (CPK)	0,063
Diabetes	-0,0019
Ejection fraction	-0,27
Sex	-0,0043
Platelets	-0,049
Serum creatinine	0,29
Serum sodium	-0,2
Smoking	-0,013
Time	-0,53

b) *Linear Discriminant Analysis*

Linear Discriminant Analysis (LDA) merupakan teknik klasifikasi sederhana tetapi *robust*. Metode ini mencari kombinasi linear dari fitur-fitur yang ada untuk memisahkan sampel ke dalam kelas yang sesuai [13]. Pada seleksi fitur dengan LDA, jumlah fitur akan direduksi dengan memaksimalkan jarak antar kelas dan meminimalkan jarak dalam kelas. Pada penelitian ini jumlah fitur terakhir hasil LDA yang digunakan adalah 1 fitur. Langkah-langkah seleksi fitur dengan *Linear Discriminant Analysis* adalah sebagai berikut:

- Pilih data yang akan digunakan.
- Hitung vektor rata-rata (*mean*) μ_i dengan mencari rata-rata untuk setiap atribut di dalam *dataset* untuk setiap kelas.

$$\mu_i = \frac{1}{N} \sum_{n=1}^{N_i} x_i^n \quad (1)$$

- Hitung matriks *covariance* kelas dan matriks *covariance* antar kelas. Dimana setiap kelas dihitung menggunakan persamaan 2.

$$X_i = \sum_{n=1}^{N_i} (x_i^n - \mu_i)(x_i^n - \mu_i)^T \quad (2)$$

Matriks *covariance* kelas X_W dan antar kelas X_B ditunjukkan pada persamaan 3 dan 4.

$$X_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \quad (3)$$

$$X_W = X_1 + X_2 \quad (4)$$

- d) Menghitung *eigen* vektor dari invers matriks *covariance*.
- e) Mengurutkan nilai yang didapat dari langkah sebelumnya secara *descending*.
- f) Reduksi vektor dengan nilai *eigen* terendah.

c) Principal Component Analysis (PCA)

Principal Component Analysis (PCA) merupakan salah satu teknik seleksi fitur (*reduksi fitur*) yang paling populer. PCA diimplementasikan menggunakan aljabar linier khususnya dekomposisi matriks seperti dekomposisi *eigen* atau SVD. Pada PCA dilakukan proyeksi data dengan m-kolom (fitur) ke dalam *subspace* dengan jumlah kolom yang lebih sedikit dari m namun tetap mempertahankan esensi data aslinya [14].

2.4. Multilayer Perceptron

Pemanfaatan metode *multilayer perceptron* sudah sering digunakan dalam berbagai bidang penelitian, salah satunya dalam bidang kesehatan. *Multilayer perceptron* adalah salah satu model jaringan syaraf tiruan. Metode ini akan memperbarui *weight* (bobot) menggunakan teknik *backpropagation*. Penelitian ini menerapkan *multilayer perceptron* sebagai metode prediksi harapan hidup pada pasien penyakit kardiovaskular. *Multilayer perceptron* terdiri dari tiga lapisan dimana lapisan pertama adalah *input layer*, lapisan kedua adalah *hidden layer*, dan yang terakhir adalah *output layer*. Pada *multilayer perceptron*, minimal terdapat satu *hidden layer* yang bisa ditambah jika dibutuhkan. Pada *hidden layer* dapat dilakukan modifikasi jumlah *neuron* atau *node*. Langkah-langkah metode *multilayer perceptron* adalah sebagai berikut [15].

- a) Berikan nilai awal bobot menggunakan bilangan acak kecil.
- b) Hitung *output* aktual dari *neuron* pada *hidden layer* dan *output layer* menggunakan fungsi aktivasi pada Persamaan 5.

$$Y_j(p) = \text{sigmoid} \left(\sum_{i=1}^n x_i(p) \cdot w_{ij}(p) - \theta_j \right) \tag{5}$$

Dimana Y_i adalah *output* aktual pada *hidden layer*, p adalah iterasi ke- p , n adalah jumlah *input neuron* j pada *hidden layer*, x_i adalah *input neuron* ke- i pada *input layer*, w_{ij} adalah bobot pada masing-masing *input* x_i yang masuk ke *neuron* j pada *hidden layer*, θ_j nilai *threshold* pada *hidden layer*, dan *sigmoid* adalah fungsi aktivasi *sigmoid*. Persamaan 6 digunakan untuk menghitung nilai aktivasi *output layer*.

$$Y_k(p) = \text{sigmoid} \left(\sum_{j=1}^m x_{jk}(p) \cdot w_{jk}(p) - \theta_k \right) \tag{6}$$

dimana Y_k adalah *output* aktual pada *output layer*, p adalah iterasi ke- p , m adalah jumlah *input neuron* k pada *output layer*, x_{jk} adalah *input neuron* ke- j pada *hidden layer*, w_{jk} adalah bobot pada masing-masing *input* x_{jk} yang masuk ke *neuron* k pada *hidden layer*, θ_k nilai *threshold* pada *output layer*, dan *sigmoid* adalah fungsi aktivasi *sigmoid*.

- c) Lakukan pembuatan model menggunakan data latih dengan menggunakan algoritma *backpropagation*.
- d) Perbarui nilai bobot mulai dari *output layer* berlanjut ke *layer* di depannya yaitu *hidden layer*. Nilai bobot baru pada *output layer* digunakan untuk menghitung bobot baru di *hidden layer*.

2.5. Evaluasi Performa

Melakukan evaluasi performa metode klasifikasi merupakan tahap penting dalam pembangunan sistem prediksi. Dengan melakukan evaluasi performa maka dapat diketahui apakah sistem berjalan dengan baik atau tidak. Penelitian ini mengevaluasi performa menggunakan beberapa metrik evaluasi yaitu *akurasi*, *precision*, *recall*, dan *f-measure* [8]. Keempat metrik ini dihitung berdasarkan *confusion matrix* pada Tabel 3.

Tabel 3. Confusion Matrix

Aktual	Prediksi	
	Positif	Negatif
Positif	<i>True Positive</i>	<i>False Positive</i>
Negatif	<i>False Negative</i>	<i>True Negative</i>

Jika data positif dan diprediksi positif akan dihitung sebagai *True Positive* (TP), tetapi jika data itu diprediksi negatif maka akan dihitung sebagai *False Negative* (FN). Jika data negatif dan diprediksi negatif akan dihitung sebagai *True Negative*, tetapi jika data tersebut diprediksi positif maka akan dihitung sebagai *False Positive* (FP).

Akurasi menunjukkan performa klasifikasi yang didefinisikan sebagai perbandingan data yang diklasifikasikan benar dengan keseluruhan data, seperti pada persamaan 7.

$$akurasi = \frac{TP+TN}{TP+FP+FN+TN} \tag{7}$$

Precision adalah perbandingan jumlah data positif yang diprediksi benar dengan keseluruhan data yang diprediksi positif, seperti pada persamaan 8.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Recall adalah perbandingan data positif yang diklasifikasikan secara benar dengan keseluruhan data positif, seperti pada persamaan 9.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

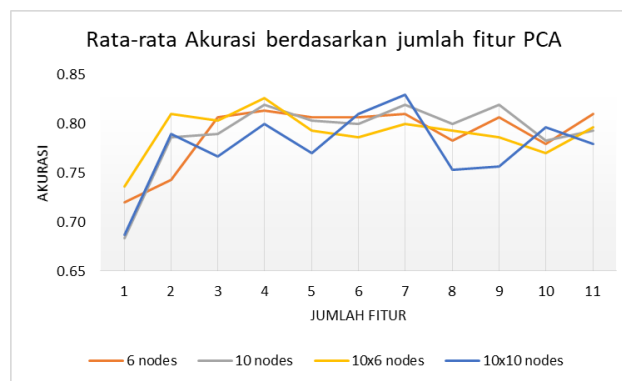
F-measure adalah perbandingan antara nilai *precision* dan *recall*, seperti pada persamaan 10. Digunakan untuk mencari model yang memiliki keseimbangan antara *precision* dan *recall*.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{10}$$

2.6. Skenario Pengujian

a) Pemilihan Jumlah Fitur pada PCA

Pada PCA akan dilakukan pemilihan jumlah fitur melalui beberapa percobaan. Pemilihan jumlah fitur dilakukan dengan melakukan percobaan dari jumlah fitur 1 hingga jumlah fitur n-1. Menggunakan *scenario* perbedaan jumlah *hidden node*, pencarian jumlah fitur maksimal pada PCA dilakukan dengan menghitung akurasi menggunakan metode klasifikasi *multilayer perceptron*. Gambar 2 menunjukkan akurasi dari MLP berdasarkan jumlah fitur yang direduksi oleh PCA.



Gambar 2. Akurasi MLP berdasarkan jumlah fitur reduksi PCA

Pada Gambar 2, terlihat fitur sejumlah 4 dan 7 menghasilkan akurasi yang cukup tinggi baik pada skenario 6 nodes, 10 nodes, 10x6 nodes (2 *hidden layer*), dan 10x10 nodes (2 *hidden layer*). Tabel 4 menunjukkan angka akurasi rata-rata. Oleh karena itu, untuk selanjutnya dipilih 4 fitur untuk skenario percobaan menggunakan seleksi fitur PCA.

Tabel 4. Angka akurasi rata-rata berdasarkan jumlah fitur PCA

Jumlah fitur	Akurasi Rata-rata
1	0.71
2	0.78
3	0.79
4	0.82
5	0.79
6	0.80
7	0.82
8	0.78
9	0.79
10	0.78
11	0.79

b) Pengujian Metode Multilayer Perceptron

Pengujian sistem prediksi harapan hidup pada pasien penyakit kardiovaskular dilakukan dengan membagi data latih dan data uji masing-masing 80% dan 20%. Pemilihan data latih dan uji diterapkan secara



proporsional. Sehingga perbandingan jumlah data antara kelas positif dan kelas negatif yang berada pada data latih dan data uji sama. Pengujian menggunakan metode *multilayer perceptron* dilakukan melalui beberapa skenario berdasarkan perbedaan jumlah *node* pada *hidden layer*, yaitu:

- a) Menggunakan seleksi fitur CBF, LDA, dan PCA dengan 1 *hidden layer*. *Node* yang digunakan berjumlah 2, 4, 6, 8, 10, dan 12. Nilai *node* lebih tinggi tidak dipakai karena tidak memberikan peningkatan signifikan terhadap performa klasifikasi.
- b) Menggunakan seleksi fitur CBF, LDA, dan PCA dengan 2 *hidden layer*. Pada *hidden layer* pertama, dipilih jumlah *node* sebanyak 10 buah. Dan pada *hidden layer* kedua, jumlah *node* bervariasi yaitu sebanyak 2, 4, 6, 8, 10, dan 12.

3. HASIL DAN PEMBAHASAN

Uji coba telah dilakukan dengan menerapkan skenario yang terdapat pada subbab 2.6. Uji coba diterapkan dengan 10 kali perulangan untuk memperoleh nilai rata-rata metrik evaluasi. Pada bagian ini akan dijelaskan nilai metrik evaluasi secara lengkap.

3.1. Perhitungan Akurasi Multilayer Perceptron dengan 1 *Hidden Layer*

Hasil uji coba menggunakan jumlah *node* 2, 4, 6, 8, 10, dan 12 dengan berbagai metode seleksi fitur ditunjukkan pada Tabel 5. Dari berbagai jumlah *node* yang diuji, rata-rata akurasi tertinggi didapat dengan jumlah *node* sejumlah 10 buah.

Tabel 5. Akurasi rata-rata berdasarkan jumlah *node* pada 1 *hidden layer*

Jumlah Layer	Akurasi Berdasarkan Seleksi Fitur			Rata-rata Akurasi
	CBF	LDA	PCA	
2	0.847	0.838	0.795	0.827
4	0.842	0.848	0.798	0.829
6	0.858	0.850	0.787	0.832
8	0.832	0.847	0.798	0.826
10	0.852	0.852	0.798	0.834
12	0.848	0.852	0.793	0.831

3.2. Perhitungan Akurasi Multilayer Perceptron dengan 2 *Hidden Layer*

Hasil uji coba menggunakan dua *hidden layer*. *Hidden layer* pertama terdapat 10 *node* dan *hidden layer* kedua dicoba dengan jumlah *node* 2, 4, 6, 8, 10, dan 12. Hasil akurasi dengan berbagai metode seleksi fitur ditunjukkan pada Tabel 6. Dari berbagai jumlah *node* yang diuji, rata-rata akurasi tertinggi didapat dengan jumlah *node* pada *hidden layer* kedua sejumlah 2 buah.

Tabel 6. Akurasi rata-rata berdasarkan jumlah *node* pada 2 *hidden layer*

Jumlah Layer	Akurasi Berdasarkan Seleksi Fitur			Rata-rata Akurasi
	CBF	LDA	PCA	
(10 x 2)	0.838	0.850	0.792	0.827

Jumlah Layer	Akurasi Berdasarkan Seleksi Fitur			Rata-rata Akurasi
(10 x 4)	0.833	0.842	0.787	0.821
(10 x 6)	0.837	0.848	0.785	0.823
(10 x 8)	0.838	0.845	0.793	0.826
(10 x 10)	0.820	0.847	0.787	0.818
(10 x 12)	0.833	0.850	0.782	0.822

3.3. Pembahasan

Berdasarkan pengujian yang telah dilakukan, metode *multilayer perceptron* memperoleh hasil akurasi yang cukup baik dengan berbagai jumlah *hidden layer* berkisar antara 80%. Dari kedua uji coba, rata-rata akurasi terbaik diperoleh dengan 1 *hidden layer* dengan jumlah 10 *node* yaitu sebesar 83.4%. Jika dilihat dari seleksi fitur, metode CBF dan LDA menghasilkan nilai akurasi yang tidak berbeda signifikan dan lebih dari 80%. Sedangkan PCA nilai akurasinya tidak dapat mencapai 80%. Metode LDA menghasilkan rata-rata akurasi yang cukup baik yaitu di angka 84.7%. Tabel 7 menunjukkan rata-rata akurasi berdasarkan metode seleksi fitur yang dipilih.

Tabel 7. Akurasi rata-rata berdasarkan metode seleksi fitur

Metode Seleksi Fitur	Akurasi Rata-rata
CBF	0.840
LDA	0.847
PCA	0.791

Karena jumlah *node* 10 pada 1 *hidden layer* menghasilkan akurasi tertinggi, maka kami mencoba melakukan uji coba kembali untuk menghitung metrik performa secara lengkap yaitu akurasi, *precision*, *recall*, dan *f-measure*. Nilai performa rata-rata dari prediksi menggunakan *multilayer perceptron* dapat dilihat pada Tabel 8.

Metode seleksi fitur CBF terlihat mampu menghasilkan rata-rata akurasi tertinggi yaitu 86,8%, sedangkan yang paling rendah adalah PCA yaitu 84,7%. *Recall* menggunakan CBF juga mampu mendapatkan skor sebesar 82,1% yang berarti mampu memprediksi data positif yang lebih sedikit juga secara akurat. Nilai *f-measure* yang menunjukkan keseimbangan antara *precision* dan *recall* juga cukup baik untuk ketiga metode seleksi fitur yaitu sebesar 79,7% untuk CBF, 76,6% untuk LDA, dan 74,9% untuk PCA.

Tabel 8. Nilai performa rata-rata

Metrik Evaluasi	CBF	LDA	PCA
<i>Accuracy</i>	0.868	0.852	0.847
<i>Precision</i>	0.777	0.767	0.780
<i>Recall</i>	0.821	0.768	0.721
<i>F-measure</i>	0.797	0.766	0.749

Berdasarkan nilai performa terbaik dari suatu uji coba, akurasi, presisi, *recall*, dan *f-measure* terbaik diperoleh dengan seleksi fitur CBF yaitu 91,7% 85% 89,5% dan 87,2%. Angka ini menunjukkan hasil yang cukup akurat

dalam prediksi harapan hidup bagi pasien kardiovaskular. Selain itu, walaupun secara rata-rata PCA memiliki akurasi terendah, namun dalam uji coba yang dilakukan f-measure terbaik menggunakan PCA sebesar 80% lebih tinggi daripada dengan LDA yaitu 77,8%.

Tabel 9. Nilai performa terbaik

Metrik Evaluasi	CBF	LDA	PCA
Accuracy	0.917	0.867	0.883
Precision	0.850	0.824	0.875
Recall	0.895	0.737	0.737
F-measure	0.872	0.778	0.800

4. SIMPULAN

Penelitian ini bertujuan untuk membuat sistem prediksi harapan hidup pasien penyakit kardiovaskular. Prediksi ini dapat menunjukkan apakah seorang pasien memiliki resiko tinggi gagal jantung atau tidak. Data yang digunakan sebagai data uji coba yaitu *open access* data dari Kaggle yang berjudul "Heart Failure Prediction". Data ini selanjutnya diproses melalui praproses untuk normalisasi. Setelah data ternormalisasi, maka akan dilanjutkan dengan seleksi fitur menggunakan tiga metode yang dijalankan secara terpisah yaitu *correlation based filter*, *linear discriminant analysis*, dan *principal component analysis*. Hasilnya akan diuji menggunakan klasifikasi *multilayer perceptron*. Berdasarkan pengujian yang dilakukan, jumlah node 10 dengan 1 *hidden layer* mendapatkan rata-rata akurasi tertinggi. Pada performa terbaiknya, nilai akurasi, *precision*, *recall*, dan f-measure yang bisa diperoleh menggunakan CBF yaitu 91,7% 85% 89,5% dan 87,2%. Berdasarkan hasil tersebut dapat disimpulkan bahwa sistem prediksi ini mampu memberikan hasil yang cukup akurat.

DAFTAR PUSTAKA

- [1] P. Kabo, Penyakit jantung koroner penyakit atau proses alamiah, Jakarta: Badan Penerbit FKUI, 2014.
- [2] K. K. RI, Pusat Data Dan Informasi Kementerian Kesehatan Ri : Situasi Kesehatan, Jakarta: Kementerian Kesehatan, 2014.
- [3] N. L. Fitriyani, M. Syafrudin, G. Alfian and J. Rhee, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System," *IEEE Access*, vol. 8, pp. 133034-133050, 2020.
- [4] P. D. S. K. I. PERKI, "Indonesian Heart Association," 2019. [Online]. Available: http://www.inaheart.org/news_and_events/news/2019/9/26/press_release_world_heart_day_perki_2019. [Accessed 22 January 2020].
- [5] F. Farzadfar, "Cardiovascular disease risk prediction models: challenges and perspectives," *The LANCET Global Health*, vol. 7, no. 10, pp. 1288-1289, 2019.

- [6] S. H. A. Aini, Y. A. Sari and A. Arwan, "Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 9, pp. 2546-2554, 2018.
- [7] O. O. Oladimeji and O. Oladimeji, "Predicting Survival of Heart Failure Patients Using Classification Algorithms," *Journal of Information Technology and Computer Engineering*, vol. 4, no. 2, pp. 90-94, 2020.
- [8] D. Christopher, R. Prabhakar and S. Hinrich, "An Introduction To Information Retrieval," *Introduction to information retrieval*, pp. 151-177, 2008.
- [9] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi and V. Padma, "Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification," in *Third International Conference on Smart Systems and Inventive Technology*, 2020.
- [10] I. H. Witten, E. Frank, M. A. Hall and C. Pal, *Data Mining: Practical Machine Learning Tools and Techniques 4th Edition*, Morgan Kaufmann, 2016.
- [11] M. Arifin, "IG-KNN untuk Prediksi Customer Churn Telekomunikasi," *Jurnal SIMETRIS*, vol. 6, pp. 1-10, 2015.
- [12] Y. Khourdifi and M. Bahaj, "Feature Selection with Fast Correlation-Based Filter for Breast Cancer Prediction and Classification Using Machine Learning Algorithms," in *International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, 2018.
- [13] M. I. Khalid, T. Alotaiby, S. A. Aldosari, S. A. Alshebeili, F. S. Y. Al-Hameed, M. Almohammed and T. Alotaibi, "Epileptic MEG Spikes Detection using Common Spatial Patterns and Linear Discriminant Analysis," *IEEE Access*, vol. 4, 2016.
- [14] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 2006.
- [15] E. P. Cynthia and E. Ismanto, "Jaringan Syaraf Tiruan Algoritma Backpropagation dalam Memprediksi Ketersediaan Komoditi Pangan Provinsi Riau," *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, vol. 2, no. 2, pp. 83-98, 2017.