
Penerapan Algoritma Naïve Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung

Duwi Cahya Putri Buani

Teknik Informatika, Universitas Nusa Mandiri

E-mail: duwi.dcp@nusamandiri.ac.id

Abstract

The heart is an important organ owned by humans, the heart has a very large role to distribute blood to all parts of the body in humans so that if there is a problem with the heart, problems will automatically occur in other organs of the body, so research on the prediction of heart failure is feasible. Nave Bayes is a classification algorithm that is easy to use but nave Bayes has weaknesses that result in poor performance, therefore feature selection is needed, one of which is to use genetic algorithms, genetic algorithms are algorithms that are able to select attributes In research, with the genetic algorithm, the attribute will be selected based on the highest weight so that the accuracy of the prediction is more optimal. Previous research using Naïve Bayes has an accuracy rate of only up to 69.60% after feature selection with genetic algorithms increases the accuracy rate to 96.67%, the difference in accuracy increases by 27.07%.

Keywords: *Genetic Algorithm, Heart Failure, Nave Bayes*

Abstarksi

Jantung merupakan organ penting yang dimiliki oleh manusia, jantung memiliki peranan yang sangat besar untuk mendistribusikan darah keseluruh bagian tubuh pada manusia sehingga jika terjadi masalah pada jantung otomatis akan terjadi masalah juga pada organ tubuh yang lain, sehingga penelitian mengenai prediksi gagal jantung layak untuk dilakukan. Naïve bayes merupakan algoritma klasifikasi yang mudah dalam penggunaannya tetapi naïve bayes memiliki kelemahan-kelemahan yang mengakibatkan performanya kurang baik, oleh sebab itu diperlukan seleksi fitur, algoritma genetika merupakan algoritma yang mampu untuk melakukan seleksi terhadap atribut-atribut dalam penelitian, dengan adanya algortima genetika maka atribut akan dipilih berdasarkan bobot yang paling tinggi agar akurasi dari prediksi lebih optimal. Penelitian sebelumnya dengan menggunakan Naïve bayes tingkat akurasinya hanya sampai dengan 69,60% setelah dilakukan seleksi fitur dengan algortitma genetika tingkat akurasi meningkat menjadi 96,67%, selisih selisih peningkatan akurasi sebesar 27,07%.

Keywords: *Algoritma Genetika, Gagal Jantung, Naïve Bayes*

1. Introduction

Jantung adalah organ vital yang dimiliki oleh manusia, jantung bertugas untuk memompa darah keseluruh tubuh, jika tidak memiliki jantung yang sehat makan selama hidup manusia akan memiliki berbagai keluhan penyakit yang akan menyerang tubuh kita[1].

Salah satu penyakit jantung adalah gagal jantung, gagal jantung merupakan suatu kegagalan otot jantung untuk memompakan darah secara memadai ke seluruh tubuh[2]. Gagal jantung merupakan salah satu penyebab kematian nomor satu di dunia, lebih banyak orang meninggal setiap tahunnya karena gagal jantung. Diperkirakan 17,9 juta orang meninggal karena gagal jantung pada tahun 2016 [3].

Gagal jantung di Indonesia terus mengalami peningkatan. Data dari data rumah sakit rujukan nasional Pusat Jantung Nasional Harapan Kita (PJNHK) menunjukkan peningkatan prevalensi gagal jantung usia di bawah 50 tahun yang dirawat inap dari 11.3% pada tahun 2011 menjadi 27.3% pada tahun 2012 [4]. Oleh sebab itu perlu adanya penelitian untuk melakukan prediksi kelangsungan hidup pasien dengan gagal jantung. Penelitian ini bukanlah penelitian yang pertama dilakukan, penelitian sebelumnya dapat dilihat pada table I berikut ini:

Tabel 1. Penelitian Terdahulu

Judul Penelitian	Metode	Akurasi
Machine learning can predict survival of patients with heart failure from serum creatinine and	Random Forests	74,00%
	Decision Tree	73,70%
	Gradient Boosting	73,80%
	Linear Regression	73,00%
	One Rule	72,90%
	Artificial Neural	68,00%

Judul Penelitian	Metode	Akurasi
ejection fraction alone	Network	
	Naïve Bayes	69,60%
	SVM Radial	69,00%
	SVM linear	68,40%
	K-Nearest Neighbors	62,40%

Sumber: [5]

Dari table I dapat dilihat bahwa metode Random Forest memiliki akurasi yang lebih tinggi dari pada metode yang lain, tingkat akurasinya mencapai 74,00%.

Pada penelitian lain, yang dilakukan oleh Rahayu dkk, dengan menggunakan metode yang sama dan dilakukan Teknik SMOTE dan Teknik resample dengan judul penelitian "Prediction Of Survival Of Heart Failure Patients Using Random Forest", tingkat akurasi untuk penggunaan SMOTE metode random forest memiliki tingkat akurasi sebesar 85,82% sedangkan penggunaan Teknik Resample metode RF atau Random Forest memiliki tingkat akurasi sebesar 94,31% [6]. Dari dua penelitian yang telah disebutkan diatas mengidentifikasi bahwa metode RF memiliki tingkat akurasi yang paling tinggi, sedangkan Naïve Bayes memiliki akurasi yang rendah, oleh karena itu didalam penelitian ini akan dilakukan peneingkatan akurasi dari naïve bayes dengan menggunakan algoritma genetika. Algoritma Genetika digunakan untuk mereduksi attribute-attribut yang kurang dominan karena metode algoritma genetika memiliki kemampuan dalam menseleksi attribut yang kurang dominan [7]. Metode Naive Bayes dapat memberikan kemudahan dalam menghitung dan dapat menentukan kemungkinan-kemungkinan yang akan terjadi, serta naïve bayes dapat menghitung dalam jumlah data yang besar [7]. Naive Bayes merupakan teknik prediksi berbasis

probabilistik sederhana yang berdasarkan pada penerapan teorema Bayes dengan asumsi independensi yang kuat. Pada metode ini terdapat kekurangan yang dapat berpengaruh pada akurasi yang disebabkan oleh fitur Naive Bayes yang tidak selalu dapat diterapkan. Untuk menangani masalah tersebut, peneliti melakukan proses seleksi fitur menggunakan Algoritma Genetika [8].

2. Materials and Methods

2.1 Dataset

Dalam melakukan penelitian pastinya akan membutuhkan dataset yang nantinya akan diolah oleh suatu algoritma, penelitian ini menggunakan dataset dari *UCI Repository Website*, dataset yang digunakan merupakan dataset publik yang di publikasi pada tahun 2020 dengan 13 atribut dapat di lihat pada table II:

Tabel 2. Atribut yang digunakan dalam penelitian

Variabel	Keterangan
Age	Umur Pasien
Anaemia	Penurunan Sel Darah Merah Atau Hemoglobin (Boolean)
High Blood Pressure	Jika Pasien Menderita Hipertensi (Boolean)
Creatinine Phosphokinase (CPK)	Tingkat Enzim CPK Dalam Darah (mcg / L)
Diabetes	Jika Pasien menderita Diabetes (Boolean)
Ejection Fraction	Persentase Darah Meninggalkan Jantung Pada Setiap Kontraksi (Persentase)
Platelets	Trombosit Dalam Darah (Kiloplatelet / mL)
Sex	Jenis Kelamin (Laki-laki atau Perempuan)
Serum Creatinine	Tingkat Kreatinin Serum Dalam Darah

Variabel	Keterangan
	(mg / dL)
Serum Sodium	Kadar Natrium Serum Dalam Darah (mEq / L)
Smoking	Jika Pasien Merokok Atau Tidak (Boolean)
Time	Periode Tindak Lanjut (Hari)
[Target] Death Event:	Jika Pasien Meninggal Dalam Masa Tindak Lanjut (Boolean)

Sumber: [5]

2.2 Metode

Metode yang digunakan dalam penelitian ini adalah metode naïve bayes dan algoritma genetika sebagai seleksi fitur. Ada banyak algoritma klasifikasi dalam data mining dan machine learning. Salah satu yang paling sederhana namun paling efektif adalah Naive Bayes classifier (NBC), NBC adalah pengklasifikasi probabilistik berdasarkan penerapan teorema Bayes dengan asumsi independensi yang kuat (naif). Singkatnya, NBC memberikan masukan ke salah satu kelas $k \in \{C_1, C_2, \dots, C_k\}$ berdasarkan beberapa properti (fitur) masukan. NBC memiliki aplikasi seperti pemfilteran spam email dan klasifikasi dokumen [9].

Sedangkan Algoritma genetika merupakan algoritma pencarian yang didasarkan pada mekanisme seleksi alamiah dan genetika alamiah. Algoritma genetika melakukan pencarian terhadap

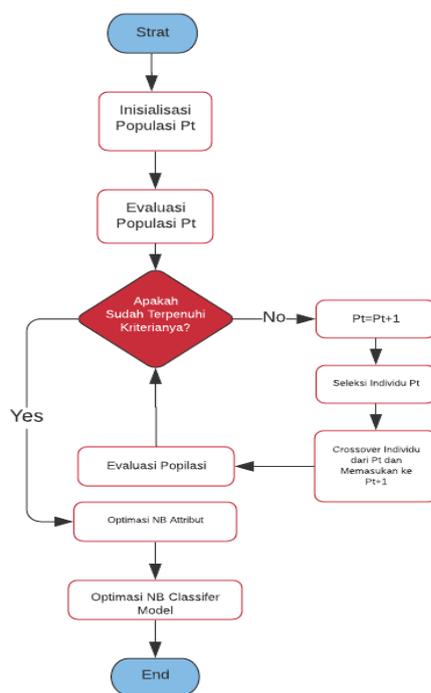
solusi optimal berupa individu dengan kualitas terbaik dari suatu populasi [10].

Seleksi fitur merupakan salah satu tahapan praproses klasifikasi. Seleksi fitur dilakukan dengan cara memilih fitur-fitur yang relevan yang mempengaruhi hasil klasifikasi. Seleksi fitur digunakan untuk mengurangi dimensi data dan fitur-fitur yang tidak relevan. Seleksi fitur digunakan untuk meningkatkan efektifitas dan efisiensi kinerja dari algoritma klasifikasi [11].

3. Results and Discussion

3.1 Proses Pengolahan Atribut menggunakan Algoritma Genetika

Tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:



Sumber: [8]

Gambar 1 Diagram Alur Naïve Bayes dan Algoritma Genetika

Gambar 1. Menunjukkan langkah

pertama dalam penggunaan Algoritma Genetika adalah inisialisasi populasi pt. Kromosom pada populasi ditentukan nilai gennya. Selanjutnya evaluasi populasi pt. Kromosom diseleksi menggunakan nilai fitness. Kromosom yang memiliki nilai fitness terbesar akan dipilih. Nilai kriteria akan terpenuhi jika jumlah generasi sudah maksimal, jika belum maksimal maka iterasi akan terus dilakukan. Kemudian berdasarkan probability crossover, kromosom terpilih akan dicrossoverkan. Dan berdasarkan probability mutation ditentukan berapa banyak gen dalam kromosom yang akan dimutasi. Setelah mencapai generasi maksimal, maka didapat kromosom dengan nilai fitness tertinggi sebagai solusi dari permasalahan seleksi atribut. Kemudian data dengan atribut yang terpilih akan ditraining dan ditesting oleh Naive Bayes [8], sehingga Naïve Bayes dapat melakukan prediksi kelangsungan hidup pasien gagal jantung.

Attribute yang terpilih dari 13 atribut adalah 7 atribut, satu atribut adalah label atau target penelitian. Atribut yang terpilih dapat dilihat pada table III:

Tabel 3. Atribut yang terpilih, yang memiliki nilai fitness tinggi

Variabel	Keterangan
Age	Umur Pasien
High Blood Pressure	Jika Pasien Menderita Hipertensi (Boolean)
Diabetes	Jika Pasien menderita Diabetes (Boolean)

Variabel	Keterangan
Ejection Fraction	Persentase Darah Meninggalkan Jantung Pada Setiap Kontraksi (Persentase)
Serum Sodium	Kadar Natrium Serum Dalam Darah (mEq / L)
Time	Periode Tindak Lanjut (Hari)
[Target] Death Event:	Jika Pasien Meninggal Dalam Masa Tindak Lanjut (Boolean)

Sumber: [8]

3.2 Penerapan Naïve Bayes dan Algoritma Genetika

Setelah terpilih atribut yang memiliki nilai fitness terbesar selanjutnya yang dilakukan adalah pengujian data dengan menggunakan Naïve bayes. Pengujian dalam penelitian ini menggunakan rapid miner dengan menghasilkan akurasi sebesar 96,67 % hasil ini lebih besar dari penelitian terdahulu dan untuk AUC (*Area Under Curve*) sebesar 0,977. Hasil penelitian dapat dilihat pada table IV:

Tabel 4. Hasil Penelitian

PZ	Max Num of Gen	TF	P Cross	Accry	AUC
10	100	1.0	1.0	94,44%	0,955
20	110	1.0	1.0	93,33%	0,956
30	120	1.0	1.0	93,33%	0,954
40	130	1.0	1.0	94,44%	0,941
50	140	1.0	1.0	94,44%	0,943
60	150	1.0	1.0	95,56%	0,973
70	160	1.0	1.0	94,44%	0,932
80	170	1.0	1.0	94,44%	0,975
90	180	1.0	1.0	95,56%	0,929
100	190	1.0	1.0	96,67%	0,977

Sumber: [8]

3.3 Conclusions

Dalam penelitian ini membuktikan bahwa Naïve Bayes performanya akan lebih

maksimal dalam melakukan prediksi jika dilakukan seleksi fitur terlebih dahulu menggunakan algoritma genetika, dimana dengan menggunakan data yang sama tanpa menggunakan algoritma genetika akurasi Naïve Bayes hanya mencapai 69,60%, setelah dilakukan seleksi fitur dengan algoritma genetika hasil akurasi meningkat menjadi 96,67%.

References:

- [1] Merdeka, "6 Fungsi Jantung pada Manusia, Ketahui Secara Teliti | merdeka.com."
- [2] Halodoc, "Penyakit Jantung - Pengertian, Gejala, Penyebab, Faktor Risiko, Diagnosis, Pencegahan, Pengobatan, Kapan Harus ke Dokter? | Halodoc.com," *HaloDoc*.
- [3] WHO, "Cardiovascular diseases (CVDs)," *12 Mei*, 2017.
- [4] www.pjnhk.go.id, "Berdamai dengan Penyakit Jantung."
- [5] 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.
- [6] S. Rahayu, J. J. Purnama, A. B. Pohan, F. S. Nugraha, S. Nurdiani, and S. Hadianti, "Prediction of survival of heart failure patients using random forest," *J. Pilar Nusa Mandiri*, vol. 16, no. 2, pp. 255–260, 2020.
- [7] S. Busono, "Optimasi Naive Bayes Menggunakan Algoritma Genetika Sebagai Seleksi Fitur Untuk Memprediksi Performa Siswa," *J. Ilm. Teknol. Inf. Asia*, vol. 14, no. 1, p. 31, 2020, doi: 10.32815/jitika.v14i1.400.
- [8] M. Y. Bakhtiar, "Klasifikasi Penelitian Dosen Menggunakan Naïve Bayes Classifier dan Algoritma Genetika," *STRING*

- (*Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 5, no. 2, p. 134, 2020, doi: 10.30998/string.v5i2.6912.
- [9] O'Reilly, "14. Naive Bayes - Data Algorithms [Book]." .
- [10] Suyanto, *Algoritma Genetika dalam MATLAB*. Yogyakarta: Andi, 2015.
- [11] H. N. Firqiani, A. Kustyo, and E. P. Giri, "Seleksi Fitur Menggunakan Fast Correlation Based Filter pada Algoritma Voting Feature Intervals 5," *J. Ilm. Ilmu Komput.*, vol. 6, no. 2, p. 245184, 2008.