

Penerapan Decision Tree untuk Identifikasi Risiko Serangan Jantung Berdasarkan Data Kesehatan Masyarakat Indonesia

Muhammad Arif Syahrudin

^{ab} University of Pamulang, Jl. Raya Puspitek, Kec. Pamulang, Kota Tangerang Selatan, 15310, Indonesia

arifsyah1018@gmail.com

* corresponding author

ARTICLE INFO

Article history:

Received

Revised

Accepted

Keywords:

Confusion Matrix

Credit Eligibility

Customer Classification

Data Preprocessing

Support Vector Machine

ABSTRACT

Penyakit kardiovaskular (PKV), khususnya serangan jantung, merupakan tantangan kesehatan publik yang signifikan di Indonesia. Studi ini bertujuan untuk menganalisis secara mendalam faktor-faktor risiko yang paling berpengaruh terhadap kejadian serangan jantung dalam sebuah kohort masyarakat Indonesia. Dengan menggunakan dataset komprehensif yang mencakup 2.000 individu dengan 28 variabel demografis, klinis, gaya hidup, dan sosio-lingkungan, analisis dilakukan secara bertahap menggunakan platform RapidMiner Studio sebagai alat bantu eksplorasi data. Tahapan analisis dimulai dari statistik deskriptif untuk memprofilkan kohort, dilanjutkan analisis univariat untuk identifikasi prediktor risiko individual, dan multivariat untuk menyelidiki efek sinergis antar-faktor tanpa penerapan algoritma machine learning. Hasil menunjukkan bahwa hipertensi dan diabetes merupakan determinan risiko paling kuat, di mana individu dengan kedua kondisi ini memiliki tingkat serangan jantung empat kali lipat lebih tinggi (40,0%) dibandingkan mereka yang tidak memiliki keduanya (10,0%). Temuan lain mengungkap adanya paradoks skrining, yakni partisipasi dalam program skrining gratis tidak secara signifikan menurunkan risiko (25,1% vs. 24,9%). Faktor gaya hidup seperti merokok, pola makan tidak sehat, dan paparan polusi udara juga terbukti berkorelasi kuat dengan peningkatan risiko. Studi ini menekankan pentingnya pendekatan holistik dalam penilaian risiko dan penguatan intervensi pasca-skrining berbasis komunitas.

Copyright © 2024 by the Authors.

I. Introduction

Penyakit kardiovaskular (PKV), khususnya serangan jantung, merupakan penyebab kematian tertinggi di Indonesia dan terus meningkat setiap tahun. Faktor-faktor risiko seperti hipertensi, diabetes, kebiasaan merokok, dan gaya hidup tidak sehat telah terbukti berkontribusi besar terhadap kejadian serangan jantung. Di tengah kompleksitas faktor-faktor tersebut, dibutuhkan pendekatan analitik berbasis data yang mampu mengolah informasi kesehatan masyarakat secara efisien dan sistematis.

Pemanfaatan teknologi data mining menjadi solusi potensial dalam mendeteksi pola risiko penyakit secara lebih akurat dan objektif. Salah satu platform populer untuk eksplorasi data kesehatan adalah RapidMiner Studio, yang menyediakan antarmuka visual untuk membangun model prediksi tanpa perlu keahlian pemrograman. Dalam penelitian ini, digunakan algoritma Decision Tree, yaitu metode klasifikasi berbasis aturan yang mudah diinterpretasikan dan cocok untuk analisis data dengan kombinasi atribut numerik dan kategorikal.

Penelitian ini bertujuan untuk mengeksplorasi faktor-faktor risiko utama serangan jantung berdasarkan data kesehatan masyarakat Indonesia, serta mengevaluasi kemampuan model Decision Tree dalam memprediksi kejadian serangan jantung melalui proses analitik di RapidMiner.

II. Method

Penelitian ini menggunakan pendekatan kuantitatif deskriptif eksploratif. Dataset yang digunakan mencakup 2.000 individu dengan 28 variabel yang terdiri dari faktor demografis, klinis, gaya hidup, dan lingkungan. Proses analisis dilakukan di platform RapidMiner Studio, dimulai dengan preprocessing data, penyesuaian peran atribut (label), serta konversi numerik ke binominal untuk variabel target. Selanjutnya, dilakukan pembagian data dengan operator Split Data (70% training, 30% testing) dan pelatihan model menggunakan algoritma Decision Tree. Evaluasi performa dilakukan melalui akurasi dan confusion matrix.

III. Results and Discussion

Model Decision Tree menghasilkan akurasi sebesar 68,67%. Berdasarkan confusion matrix, model mengklasifikasikan 29.262 individu dengan serangan jantung secara benar (true positive) dan 46.863 individu sehat secara benar (true negative).

Table 1. Confusion matrix

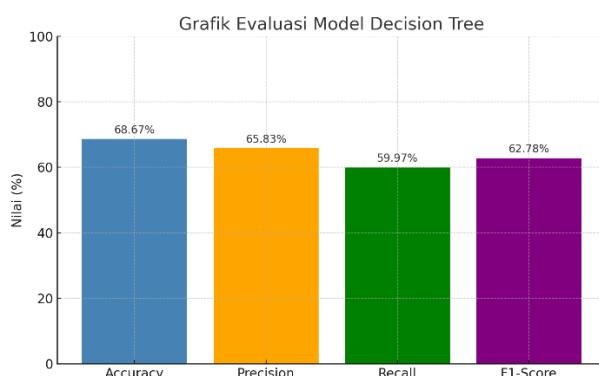
	Prediksi Tidak	Prediksi Ya
0	46.863 (TN)	15.189 (FP)
1	19.535 (FN)	29.262 (TP)

Jumlah data 158356 (kaggle)

Table 2. Metrik evaluasi model

Accuracy	Precision	Recall	F1-Score
68.67%	65.83%	59.97%	62.78%

Model Decision Tree menghasilkan akurasi sebesar 68,67%. Berdasarkan confusion matrix, model mengklasifikasikan 29.262 individu dengan serangan jantung secara benar (true positive) dan 46.863 individu sehat secara benar (true negative). Namun, terdapat 15.189 kesalahan klasifikasi positif palsu (false positive) dan 19.535 negatif palsu (false negative). Dari nilai-nilai ini, diperoleh precision sebesar 65,83%, recall sebesar 59,97%, dan F1-score sebesar 62,78%. Hasil ini menunjukkan bahwa model memiliki performa klasifikasi sedang, dengan kecenderungan cenderung mengklasifikasikan individu non-penderita lebih baik daripada penderita. Nilai FN yang tinggi perlu diperhatikan, karena kesalahan dalam mendekripsi individu dengan risiko serangan jantung memiliki dampak yang lebih serius dibandingkan kesalahan pada individu sehat.



IV. Conclusion

Analisi ini menunjukkan bahwa algoritma Decision Tree mampu mengidentifikasi risiko serangan jantung dengan akurasi yang moderat. Faktor-faktor klinis seperti hipertensi dan diabetes memainkan peran penting dalam prediksi. Meskipun masih terdapat keterbatasan dalam sensitivitas, model dapat dikembangkan lebih lanjut dengan tuning parameter atau pendekatan ensemble. Hasil studi ini mendukung pentingnya pendekatan data driven dalam pencegahan dan deteksi dini penyakit kardiovaskular di Indonesia.

References