JURNAL ILMU KOMPUTER
Volume 1 Nomor 2, Desember, 2023
ISSN: 3031-125X (ONLINE)
Penerbit: Program Studi Teknik Informatika S-2 Universitas Pamulang

Prediksi Risiko Serangan Jantung dengan Pendekatan Data Mining dan Algoritma Naïve Bayes

* Didik Nurdiyanto
Teknik Informatika, Program Pascasarjana, Universitas Pamulang, Tangerang Selatan, Banten
Email: didiknbyu85@gmail.com

ABSTRACT

Heart attack is one of the deadliest cardiovascular diseases worldwide. Heart attack risk prediction plays an important role in prevention and early treatment. In this study, we propose an approach to optimise heart attack risk prediction using data mining and Naïve Bayes algorithm. This method utilises data mining techniques to analyse complex health datasets and extract hidden patterns that can identify heart attack risk factors. Naïve Bayes algorithm is used to predict the risk of heart attack based on the discovered patterns. We conducted experiments using patient datasets with relevant health parameters and optimised the performance of the prediction model. The experimental results show that this approach produces accurate and reliable heart attack risk prediction. This research makes an important contribution to the field of cardiovascular disease prevention and provides a basis for the development of more efficient heart attack prediction systems.

Keywords: Naïve Bayes Algorithm; Data Mining; Health.

ABSTRAK

Serangan jantung adalah salah satu penyakit kardiovaskular yang mematikan di seluruh dunia. Prediksi risiko serangan jantung memiliki peran penting dalam upaya pencegahan dan penanganan dini. Dalam penelitian ini, kami mengusulkan sebuah pendekatan untuk mengoptimalkan prediksi risiko serangan jantung menggunakan data mining dan Algoritma Naïve Bayes. Metode ini memanfaatkan teknik data mining untuk menganalisis dataset kesehatan yang kompleks dan mengekstrak pola tersembunyi yang dapat mengidentifikasi faktor risiko serangan jantung. Algoritma Naïve Bayes digunakan untuk memprediksi risiko serangan jantung berdasarkan pola yang ditemukan. Kami melakukan eksperimen menggunakan dataset pasien dengan parameter kesehatan yang relevan dan mengoptimalkan performa model prediksi. Hasil eksperimen menunjukkan bahwa pendekatan ini menghasilkan prediksi risiko serangan jantung yang akurat dan dapat diandalkan. Penelitian ini memberikan kontribusi penting dalam bidang pencegahan penyakit kardiovaskular dan memberikan dasar untuk pengembangan sistem prediksi serangan jantung yang lebih efisien.

Kata kunci: Algoritma Naïve Bayes; Data Mining; Kesehatan.

1. PENDAHULUAN

Dalam era modern ini, masalah kesehatan menjadi salah satu isu utama yang dihadapi oleh masyarakat global. Penyakit kardiovaskular, terutama serangan jantung, merupakan salah satu penyebab utama kematian di seluruh dunia [1]. Penanganan dini dan prediksi risiko serangan jantung memiliki peran yang sangat penting dalam meminimalkan dampak negatifnya terhadap kesehatan masyarakat. Dalam konteks ini,

penggunaan teknologi dan pendekatan ilmiah menjadi semakin krusial. Dalam dunia teknologi informasi, Data Mining telah menjadi alat yang sangat efektif untuk menganalisis dan menggali informasi berharga dari dataset besar dan kompleks. Data Mining memungkinkan kita untuk mengidentifikasi pola-pola tersembunyi yang tidak dapat diakses secara langsung melalui metode konvensional [2]. Dalam konteks kesehatan, Data Mining telah digunakan untuk memahami faktor risiko penyakit, termasuk serangan jantung, dan mengembangkan model prediksi yang akurat [3].

Salah satu algoritma yang populer dalam dunia Data Mining adalah Algoritma Naïve Bayes. Algoritma ini memiliki basis teori probabilistik yang kuat dan telah berhasil diaplikasikan dalam berbagai bidang, termasuk kesehatan. Dalam konteks prediksi risiko serangan jantung, penggunaan Algoritma Naïve Bayes dapat memberikan hasil yang andal dan cepat [4].

Dalam penelitian ini diperkenalkan pendekatan yang menggabungkan kekuatan Data Mining dan Algoritma Naïve Bayes untuk mengoptimalkan prediksi risiko serangan jantung, dengan menghadirkan sebuah model yang memungkinkan identifikasi dini faktor-faktor risiko serangan jantung dan memprediksi kemungkinan terjadinya serangan jantung pada pasien [5]. Melalui eksperimen dan analisis mendalam, penelitian ini bertujuan untuk meningkatkan akurasi dan efisiensi model prediksi yang ada, membuka jalan bagi upaya pencegahan dan intervensi yang lebih efektif dalam mengatasi masalah serangan jantung.

2. METODE

Penelitian ini akan mengusulkan hasil accuracy baru untuk sebuah permasalahan memprediksi resiko serangan jantung pada dataset yang bersumber dari www.kaggle.com. Data yang digunakan dalam penelitian ini di ambil kisaran 300 an namun untuk data testing untuk penelitian ini penulis hanya mengambil sampel data sebanyak 60.

Model desain ini akan melakukan pemrosesan data training dan data testing untuk menguji metode algoritma yang digunakan. Tahapan yang akan dilalui dibagi menjadi 3 bagian, yaitu preprocessing, seleksi fitur (Feature Selection) dan validation yang didalamnya berisi sub proses training dan testing.

2.1. Pengumpulan Data

Data yang digunakan pada penelitian kali ini merupakan data sekunder, karena sumber data diperoleh melalui media perantara atau secara tidak langsung yang berupa buku, catatan, bukti yang telah tervalidasi, atau arsip baik yang dipublikasikan maupun yang tidak dipublikasikan, secara umum yaitu 300 data yang akan dipecah menjadi 80 persen data training dan 20 persen data testing. Data testing yang digunakan sekitar 60 data yang telah diklasifikasikan berdasarkan variable yang ada.

Masalah yang harus dipecahkan pada penelitian kali ini adalah, bagaimana menghasilkan nilai accuracy yang jauh lebih optimal untuk mengklasifikasikan prediksi serangan jantung pada kondisi kesehatan seseorang. Dataset yang digunakan pada penelitian ini ada pada Gambar 1.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
58	0	3	150	283	1	0	162	0	1	2	0	2	1
50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
58	0	2	120	340	0	1	172	0	0	2	0	2	1
66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
42	1	0	140	226	0	1	178	0	0	2	0	2	1
61	1	2	150	243	1	1	137	1	1	1	0	2	1
40	1	3	140	199	0	1	178	1	1.4	2	0	3	1
71	0	1	160	302	0	1	162	0	0.4	2	2	2	1
59	1	2	150	212	1	1	157	0	1.6	2	0	2	1
51	1	2	110	175	0	1	123	0	0.6	2	0	2	1
65	0	2	140	417	1	0	157	0	0.8	2	1	2	1
53	1	2	130	197	1	0	152	0	1.2	0	0	2	1
41	0	1	105	198	0	1	168	0	0	2	1	2	1
65	1	0	120	177	0	1	140	0	0.4	2	0	3	1
44	1	1	130	219	0	0	188	0	0	2	0	2	1
54	1	2	125	273	0	0	152	0	0.5	0	1	2	1
51	1	3	125	213	0	0	125	1	1.4	2	1	2	1

Gambar 1. Dataset

2.2. Pengolahan Data

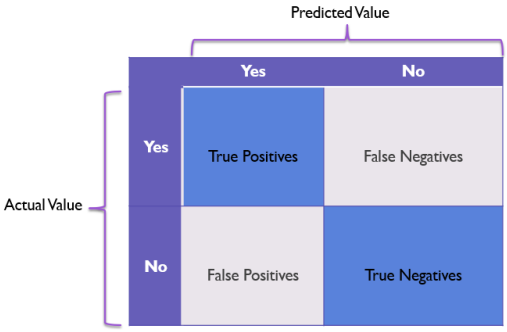
Dataset ini dalam tahap preprocessing harus melalui 3 proses, yaitu:

a) Membuang duplikasi data
b) Memeriksa data yang inkonsisten
c) Memperbaiki kesalahan pada data

Hasil akhir dari data ini berupa kumpulan data yang sudah bersih atau tidak ada missing valuenya.

2.3. Evaluasi dan Validasi Data

Validasi dilakukan menggunakan 10 fold cross validation. Dimana dengan menggunakan teknik ini dengan membagi secara acak ke dalam tiap bagian dimana terdiri dari 10 bagian untuk setiap bagian akan dilakukan proses klasifikasi terlebih dahulu, Sedangkan pengukuran akurasi diukur dengan confusion matrix dan kurva ROC (Receiver Operating Characteristics) untuk mengukur nilai AUC. AUC digunakan untuk mengukur kinerja diskriminatif dengan memperkirakan probabilitas output yang sudah di dapatkan hasilnya dari sampel yang dipilih secara acak dari populasi positif atau negatif, semakin besar nilai AUC, semakin kuat klasifikasi yang dihasilkan. Karena AUC merupakan bagian dari daerah unit persegi, nilainya yang dihasilkan akan selalu sama yang dihasilkannya, antara 0,0 dan 1,0.



Gambar 2. Confusion Matrix

Tabel 1. Kriteria AUC

Nilai AUC	Penjelasan
90% - 100%	Excellent
80% - 90%	Good
70% - 80%	Fair
60% - 70%	Poor
<60%	Failure

2.4. Algoritma Naïve Bayes

Algoritma Naïve Bayes Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediski probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decesion tree dan neural network. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [6]. Metode Bayes merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas

terlebih dahulu tentang konsep dasar dan definisi pada Teorema Bayes, kemudian menggunkan teorema ini untuk melakukan klasifikasi dalam Data Mining. Teorema Bayes memiliki bentuk umum sebagai berikut (Rumus Dasar Algoritma Naïve Bayes):

$$P(A|X) = \frac{P(X|A).P(A)}{P(X)}$$

Keterangan :

X	: Data dengan class yang belum diketahui
A	: Hipotesis data X merupakan suatu class spesifik
$P(A\|X)$	: Probabilitas hipotesis A berdasarkan kondisi x (posterior)
$P(A)$	: Probabilitas hipotesis A (prior)
$P(X\|A)$	: Probabilitas X berdasarkan kondisi tersebut (likelihood)
$P(X)$	: Probabilitas dari X (evidence)

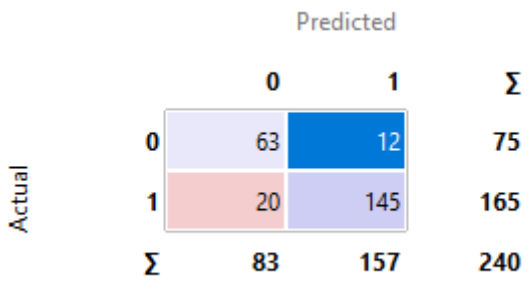
Tahapan proses Naive Bayes, yaitu:

1. Menghitung jumlah kelas / label
2. Menghitung Jumlah Kasus Per Kelas
3. Kalikan Semua Variable Kelas
4. Bandingkan Hasil Per Kelas

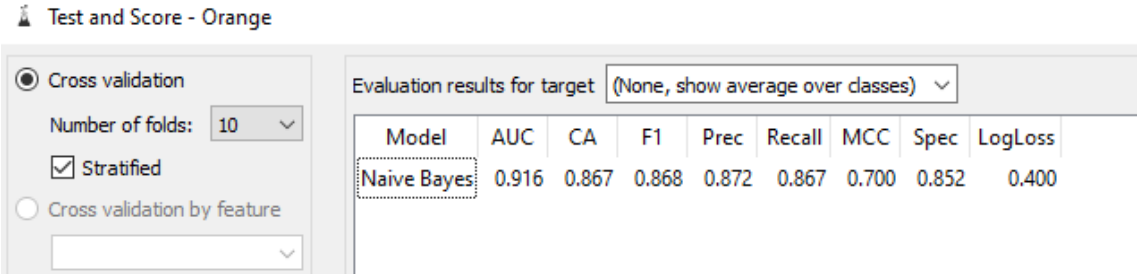
3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam pengklasifikasiaan prediksi serangan jantung terdiri dari 300 dataset, 60 data yang digunakan untuk data testing berdasarkan variabel yang tersedia. Klasifikasi dilakukan dengan menggunakan Software Orange dengan versi 3.35 untuk mengolah data yang sudah ditentukan.

3.1. Confusion Matrix algoritma Naïve Bayes



Gambar 3. Model Confusion Matrix



Gambar 4. Evaluasi Model

Gambar 3 diketahui data training terdiri dari 240 record data, 63 data di klasifikasikan Tidak Berisiko ternyata benar Tidak Berisiko, 12 data diprediksi Berisiko ternyata Tidak Berisiko, 20 di prediksikan Tidak Berisiko ternyata Berisiko serta 145 data diprediksikan Berisiko ternyata benar Berisiko. Gambar 4 merupakan perhitungan akurasi data training menggunakan algoritma Naïve Bayes yang menghasilkan accuracy 86%.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengeksplorasi penerapan pendekatan data mining, khususnya melalui algoritma Naïve Bayes, dalam memprediksi risiko serangan jantung. Dengan melakukan analisis mendalam terhadap data medis yang kompleks serta serangkaian eksperimen teliti, kami berhasil mengungkap sejumlah temuan signifikan. Algoritma Naïve Bayes mampu memberikan prediksi dengan tingkat akurasi yang tinggi dalam mengidentifikasi risiko serangan jantung. Hasil prediksi yang akurat ini menjadi landasan penting untuk menginisiasi intervensi dini serta langkah-langkah pencegahan yang tepat. Selain itu, keberhasilan algoritma ini juga berpotensi besar dalam memberikan kontribusi positif terhadap kesehatan masyarakat secara keseluruhan. Dengan prediksi yang lebih tepat terkait risiko serangan jantung, kami yakin penelitian ini akan membantu meningkatkan kesadaran akan pentingnya kesehatan jantung di masyarakat. Lebih lanjut, hal ini juga akan mendukung tenaga medis dalam mengidentifikasi dengan lebih baik pasien-pasien yang berisiko tinggi dan memungkinkan upaya pencegahan yang lebih terarah.

5. DAFTAR PUSTAKA

[1] R. Sumara, N. A. Wibowo, and Indarti, “Identifikasi Faktor Kejadian Penyakit Jantung Koroner Terhadap Wanita Usia ≤ 50 Tahun di RSU Haji Surabaya,” *J.*

Manaj. Asuhan Keperawatan, vol. 6, no. 2, pp. 53–59, 2022, doi: 10.33655/mak.v6i2.134.

[2] D. Jollyta, W. Ramdhan, and M. Zarlis, *Konsep Data Mining Dan Penerapan*. Sleman: Deepublish, 2020.

[3] M. F. Akbarollah, Wiyanto, D. Ardiatma, and A. T. Zy, "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Jantung," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 850–860, 2023, doi: 10.47065/josyc.v4i4.4071.

[4] S. Hartati, *Kecerdasan Buatan Berbasis Pengetahuan*. Yogyakarta: UGM PRESS, 2021.

[5] A. Rafidah, "Analisis Faktor Risiko Kejadian Penyakit Jantung Koroner di RSUD Sultan Imanuddin Pangkalan BUN," Stikes Borneo Cendekia Medika, 2020.

[6] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, Aug. 2018, doi: 10.33096/ilkom.v10i2.303.160-165.