

Klasifikasi Penyakit Jantung Menggunakan Metode Naïve Bayes

Ayu Novita Sari¹, Selly Alfionita²

^{1,2}Program Studi Informatika, Sekolah Tinggi Teknik Pati
Email : ayun30739@gmail.com, sellyalfio18@gmail.com

Artikel Info

Kata kunci

Penyakit jantung
Atribut
Naïve Bayes

ABSTRAK

Penyakit jantung merupakan salah satu penyakit yang banyak menimbulkan kematian yang paling banyak, tercatat 12 juta jiwa kematian setiap tahunnya di seluruh dunia. Hal ini yang terkadang memerlukan pemeriksaan sejak dini untuk mengetahui penyakit jantung tersebut. Tetapi proses diagnose merupakan hal yang cukup menantang dikarenakan keterkaitan yang cukup kompleks diantara atribut – atribut penyakit jantung. Sehingga memerlukan untuk mengetahui atribut utama yang digunakan sebagai proses pengambilan keputusan atau proses klasifikasi pada penyakit jantung. Pada penelitian ini menggunakan 14 atribut di dalamnya. Maka diperlukan penelitian agar mempersingkat klasifikasi pada penyakit jantung. Pada proses klasifikasi menggunakan Metode Naïve Bayes untuk meningkatkan nilai akurasi pada klasifikasi penyakit jantung. Tingkat Akurasi yang didapatkan 85,33%.

Penulis Korespondensi :

Ayu Novita Sari,
Program Studi Informatika
Sekolah Tinggi Teknik Pati
Email: ayun30739@gmail.com

1. PENDAHULUAN

Penyakit jantung koroner bisa disebut penyakit kardiovaskular, yaitu menyempit atau tersumbat pembuluh darah yang menimbulkan serangan jantung, disertai nyeri dada atau stroke [1]. Penyakit jantung salah satu penyakit yang memiliki angka kematian dengan tingkat yang tinggi lebih dari 12 juta jiwa kematian yang terjadi pada seluruh dunia dikarenakan penyakit jantung ini [2]. Diagnosa secara dini sangat penting untuk dilakukan, karena hal tersebut sangat menantang dikarenakan saling ketergantungan yang kompleks dari beberapa faktor atribut. Permasalahan yang sering kali dihadapi adalah kurangnya akurasi pada proses klasifikasi [3]. Penyakit jantung adalah penyakit yang sangat banyak memakan korban jiwa di dunia. Penyakit jantung ini mengancam semua jenis usia.

Machine learning adalah subdivisi yang muncul dari kecerdasan buatan. Fokus utamanya adalah merancang sistem, memungkinkan mereka untuk belajar dan membuat prediksi berdasarkan pengalaman [4]. Artificial Neural adalah bagian dari pembelajaran mesin yang menyerupai seperti cara kerja otak manusia dalam menangkap informasi dan memberikan respon pada saat yang bersamaan, karena Artificial Neural dibuat untuk mempermudah manusia diberi kecerdasan buatan [5].

Penyakit jantung yang merupakan salah satu jenis penyakit mematikan dengan melakukan analisa terhadap gejala-gejala dan data kesehatan tubuh dari penyakit jantung yang diperoleh dari data public dengan tujuan untuk mempresentasikan dengan data kesehatan yang telah dianalisa apakah sesuai antara prediksi yang dihasilkan dari perhitungan dengan menggunakan machine learning berdasarkan algoritma Neural Network dan Algoritma Naïve Bayes [6]. Analisis data tersebut dapat dilakukan dalam prediksi penyakit jantung menggunakan teknik pembelajaran mesin yang sangat efisien bagi tenaga medis [7]. Beberapa penelitian terdahulu mengambil data UCI telah melakukan penelitian tentang prediksi penyakit jantung dengan dataset publik yang bersumber dari UCI Machine Learning dengan menggunakan berbagai metode prediksi pada Penelitian [8]. Penyakit jantung atau yang disebut juga koroner adalah penyakit dengan kondisi dimana adanya timbunan lemak di pembuluh darah arteri koroner pada jantung dan adanya hambatan aliran

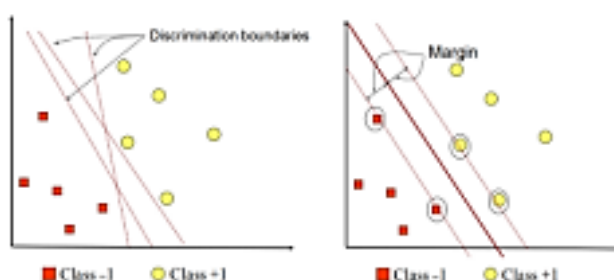
darah menuju jantung [10]. Klasifikasi Naive Bayes adalah salah satu metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik. Berdasarkan teorema Bayes, yang mengasumsikan bahwa setiap fitur memberikan kontribusi yang independen dan setara kelas target [12]. Naive Bayes adalah model pembelajaran mesin yang memanfaatkan perhitungan klasifikasi probabilitas dan statistik. Naive Bayes didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network [13].

2. METODE

Penelitian ini menggunakan metode pembelajaran mesin dalam prediksi penyakit jantung yang terdiri dari Support Vector Machine (SVM), Neural Network (NN).

A. Support Vector Machine (SVM)

Support Vector Machine adalah metode yang mempelajari untuk melakukan klasifikasi. SVM dapat digunakan pada klasifikasi linier atau non linier [7]. Hyperplane yang paling baik dapat ditemukan dengan mengukur margin hyperplane dan mencari titik maksimalnya. Pemisahan kelas-kelas oleh SVM pada gambar.



Gambar 1. Ilustrasi SVM

Ada beberapa pattern yang terkelompok dalam 2 kelas -1 (merah) dan +1 (kuning). Terdapat 2 kasus dalam memisahkan kelas oleh Hyperplane yaitu kelas yang dapat dipisahkan secara sempurna disebut dengan SVM linier dan kelas yang tidak dapat dipisahkan dengan sempurna disebut SVM non linier.

Pada dasarnya SVM non linier merupakan solusi dari masalah SVM linier dengan melakukan fungsi kernel pada ruang fitur dimensi tinggi. Definisi persamaan SVM linier dan non linear dapat dilihat pada Tabel 1.

Tabel 1. Definisi Persamaan SVM Linier dan Non Linier

Sifat SVM	Jenis Kernel	Definisi Rumus
SVM Linier	Linier	$K(x,y) = x.y$
SVM Non Linier	Polynomial	$K(x,y) = (x.y + 1)^p$
	Gaussian RBF	$K(x,y) = e^{-\ x-y\ ^2/2\sigma^2}$
	Sigmoid	$(x,y) = \tanh(Kx.y - \delta)$

Keterangan:

p : pangkat (degree of)

σ : nilai sigma

δ : nilai delta

B. Neural Network (NN)

Neural Network (NN) atau bias juga disebut Artificial Neural Network (ANN) adalah algoritma Machine Learning yang diciptakan dengan meniru jaringan neuron pada otak manusia dengan tujuan agar mesin dapat belajar mirip dengan cara belajar otak manusia, algoritma ini tergolong algoritma paling cerdas sebagai dasar pembuatan AI.

Prediksi terhadap kejadian yang akan datang setelah mempelajari kejadian dimasa lalu atau kejadian yang pernah dialami sebelumnya. Peramalan juga bias digunakan dalam aktivitas bisnis dimana dapat memperkirakan jumlah penjualan dan penggunaan produk di periode yang akan datang, sehingga produk dapat dibuat dalam kuantitas yang tepat sesuai dengan hasil peramalan [14].

Penelitian ini merupakan penelitian eksperimen dengan menggunakan dataset sekunder yang telah disediakan oleh Center for Machine Learning and Intelligent Systems [4]. Penelitian ini menggunakan 303 dataset sekunder yang terdiri dari 165 record atribut label sakit dan 138 record atribut label sehat, dengan 14 atribut (features), diantaranya:

Tabel 2. Informasi Atribut

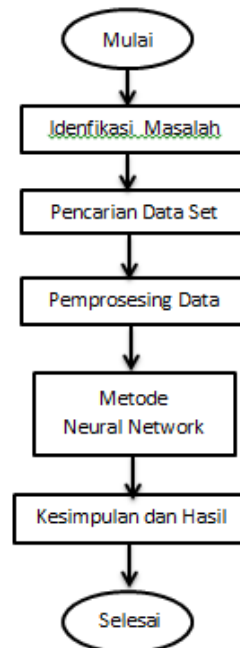
No	Atribut	Keterangan
1	Age (Usia)	-
2	Sex	0 = Female 1 = Male
3	ChestPain	Typical Asymptotic Nonanginal Nontypical
4	RestBP	Rest Blood Pressure
5	Chol	Serum Kolesterol dalam mg/dl
6	Fbs	Fasting blood sugar > 120 mg/dl 0 = False 1 = True
7	Restecg	Resting Electrocardiographic result
8	MaxHR	Maximum heart rate achieved
9	ExAng	Exercise induced angine 0 = no 1 = yes
10	Oldpeak	ST depression induced by exercise relative to rest
11	Slope	Slope of the peak exercise ST segment
12	Ca	Number of major vessels colored by flourosopy (0-3)
13	Thal	3 = normal 6 = fixed defect 7 = reversible defect
14	Target	AHD – Diagnosis heart disease 0 = no 1 = yes

3. PEMBAHASAN HASIL

Penelitian ini menggunakan dataset yang berasal dari UCI Repository Machine Learning yang terdiri dari 14 atribut jumlah dataset 303 pada data penyakit jantung. Penelitian ini menggunakan Metode Naïve Bayes untuk menghasilkan akurasi yang lebih baik dan akurat. Dataset terdiri dari 14 atribut diantaranya age, sex, cp, restbp, chol, fbs, restecg, maxhr, exang, oldpeak, slope, ca, thal dan target.

Tabel 3. Data Set Penyakit Jantung

Metodologi penelitian merupakan tahapan-tahapan yang sistematis dilakukan pada penelitian ini sehingga penelitian ini terarah dengan baik. Berikut adalah metodologi penelitian yang dilakukan.



Gambar 2. Metodologi Penelitian

3.1 Hasil Pengujian

Pada awal penelitian dalam pengujian, terlebih dahulu untuk melakukan klasifikasi pada dataset menggunakan Neural Network. Setelah dilakukan klasifikasi lalu dilakukan proses selanjutnya menggunakan metode Naïve Bayes. Kombinasi atribut pada Naïve Bayes dapat melihat dari nilai akurasi klasifikasi penyakit jantung. Adapun hasil dari proses klasifikasi dengan algoritma Neural Network dilakukan melihat akurasi ini:

accuracy: 78.57% +/- 5.69% (mikro: 78.55%)			
	true 1.0	true 0.0	class precision
pred. 1.0	132	32	80.49%
pred. 0.0	33	106	76.26%
class recall	80.00%	76.81%	

Gambar 3. Tingkat Akurasi menggunakan Neural Network

Pada Gambar 3 terdapat tingkat akurasi pada kategori yang berbeda –beda, yaitu Precision, Recall dan Accuracy. Precision adalah ketetapan tingkat akurasi antara informasi yang tersedia pada dataset dapat memberikan hasil akurasi. Accuracy adalah tingkat kedekatan antara nilai yang tepat. Dari hasil klasifikasi tanpa dilakukan tingkat akurasi sebesar 78.55%.

Kemudian setelah dilakukan klasifikasi dan mengukur tingkat akurasi pada dataset menggunakan metode Naïve Bayes dengan menggunakan 14 atribut dapat meningkatkan nilai akurasi untuk klasifikasi penyakit jantung.

accuracy: 85.38% +/- 6.27% (mikro: 85.33%)			
	true 1.0	true 2.0	class precision
pred. 1.0	10	19	34.48%
pred. 2.0	8	147	94.84%
class recall	55.56%	88.55%	

Gambar 4. Metode Naïve Bayes

Dari hasil pada Gambar 4 terdapat persentase yang berbeda seperti precision, recall, dan accuracy. Menggunakan Metode Naïve Bayes mengikat nilai akurasi sebesar 85,33%.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan terhadap dataset UCI mengenai klasifikasi penyakit jantung menggunakan Metode Neural Network memiliki kemampuan dalam meningkatkan akurasi pada algoritma Naïve Bayes untuk menentukan klasifikasi resiko Penyakit Jantung dengan hasil nilai akurasi terbaik sebesar 85.33% presisi sebesar 94,84% dan recall sebesar 88,55% menggunakan data dari UCI dataset. Metode Neural Network yang mendapatkan hasil lebih rendah yaitu 78,55%.

REFERENCES

- [1] F. Babič, J. Olejár, Z. Vantová and J. Paralič, "Predictive and Descriptive Analysis for Heart Disease Diagnosis," *Federated Conference on Computer Science and Information Systems*, vol. 11, pp. 155-163, 2017.
- [2] D. B. Umadevidan M. Snehapriya, "A Survey on Prediction of Heart Disease Using Data Mining Techniques," *International Journal of Science and Research (IJSR)*, vol. 6, no. 4, pp. 2228 - 2232, 2017.
- [3] A. Jain, M. Ahirwardan R. Pandey, "A Review on Intutive Prediction of Heart Disease Using Data Mining Techniques," *International Journal of Computer Sciences and Engineering*, vol. 7, no. 7, pp. 109 - 113, 2019.
- [4] Al Hafiz, D. M., Amaly, K., Jonathan, J., Rachmatullah, M. T., & Rosidi, R. (2021). SISTEM PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN METODE NAIVE BAYES. *Jurnal Rekayasa Elektro Sriwijaya*, 2(2), 151-157.
- [5] Handayani, F. Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 7(3), 329-334.
- [6] Riani, A., Susianto, Y., & Rahman, N. (2019). Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes. *Journal of Innovation Information Technology and Application (JINITA)*, 1(01), 25-34.
- [7] R. Katarya dan S. Kumar, "Machine Learning Techniques for Heart Disease Prediction : A Comparative Study and Analysis," *Health Technol. (Berl.)*, no. 123456789, 2020.
- [8] R. Kannandan V. Vasanthi, "Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the Heart Disease," *Spinger*, 2019.
- [9] P. W. Wahyuni, C. H. Rosjidi, S. Nurhidayat, "Identifikasi usia Sebagai Faktor Risiko Penyakit Jantung Koroner Pada Perempuan Di Poli Jantung RSUD Dr. HarjonoPonorogo", *Health Science Journal Vol 3, No.1 ISSN 2598-1196(Online)*, April, 2019.
- [10] A. Janosi, W. Steinbrunn, M. Pfisterer, R. Detrano, "Heart Disease Data Set", Center for Machine Learning and Intelligent Systems, Tersedia di: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>.
- [11] Samosir, A., Hasibuan, M. S., Justino, W. E., & Hariyono, T. (2021, September). Komparasi Algoritma Random Forest, Naïve Bayes dan K-Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung. In *Prosiding Seminar Nasional Darmajaya* (Vol. 1, pp. 214-222).
- [12] Nawawi, H. M., Purnama, J. J., & Hikmah, A. B. (2019). KOMPARASI ALGORITMA NEURAL NETWORK DAN NAÏVE BAYES UNTUK MEMPREDIKSI PENYAKIT JANTUNG. *Jurnal PILAR Nusa Mandiri*.