

ALGORITMA KLASIFIKASI DATA MINING NAÏVE BAYES BERBASIS *PARTICLE SWARM* *OPTIMIZATION* UNTUK DETEKSI PENYAKIT JANTUNG

Nur Aeni Widiastuti¹, Stefanus Santosa², Catur Supriyanto³

^{1,2,3} Pasca Sarjana Teknik Informatika, Universitas Dian Nuswantoro, Semarang 50131

¹diaz_tp05@yahoo.com

²stefanus@research.dinus.ac.id

³catur@research.dinus.ac.id

Abstrak: Data mining sering disebut *Knowledge Discovery in Database* (KDD). Data mining biasanya digunakan untuk memperbaiki pengambilan keputusan di masa yang akan datang berdasarkan informasi yang diperoleh dari masa lalu. Misalnya untuk prediksi, estimasi, asosiasi, *clustering*, dan deskripsi. Sekumpulan data yang ada di laboratorium klinik belum difungsikan secara efektif dan hanya difungsikan sebagai arsip untuk riwayat penyakit pasien. Jantung merupakan pembunuh nomor satu di dunia. Kurangnya aliran darah dan oksigen ke jantung bisa menyebabkan penyakit jantung. Pada penelitian ini akan membandingkan algoritma klasifikasi data mining Naive Bayes Berbasis PSO untuk deteksi penyakit jantung. Pengukuran dengan Naives Bayes menghasilkan akurasi 82.14%, sementara dengan Naives Bayes Berbasis *Particle Swarm Optimization* akurasi meningkat menjadi 92.86%. Tingkat akurasi dibandingkan dengan hasil laboratorium.

Kata Kunci: Data Mining, Penyakit Jantung, Naive Bayes, *Particle Swarm Optimization*.

Abstract: *Data mining is often called knowledge discovery in databases (KDD). Data mining is usually used to improve decision making in the future based on information obtained from the past. For example, for prediction, estimation, association, clustering, and description. A set of data in the clinical laboratory has not functioned effectively and simply functioned as an archive for the history of the patient's disease. The heart is the number one killer in the world. Lack of blood flow and oxygen to the heart can cause heart disease. This research will compare the data mining classification algorithm naive Bayes-based PSO for detection of heart disease. Measurements with Naives Bayes produces 82.14% accuracy, while with Naives Particle Swarm Optimization Based Bayes accuracy increased to 92.86%. The level of accuracy compared to laboratory results.*

Keywords: *Data Mining, Heart Disease, Naive Bayes, Particle Swarm Optimization.*

I. PENDAHULUAN

Penyakit jantung merupakan pembunuh nomor satu di dunia. Setiap tahun, lebih dari 2 juta orang Amerika meninggal karena penyakit jantung/stroke [1]. Penyakit jantung merupakan gangguan yang

terjadi pada sistem pembuluh darah besar. Sehingga menyebabkan jantung dan peredaran darah tidak bekerja sebagaimana mestinya [2]. Penyakit jantung sering dikenal dengan “*Sudden Death*”.

Tingginya faktor kematian akibat penyakit jantung dapat dicegah dan ditekan faktor resikonya. Kurangnya pengetahuan masyarakat tentang gejala penyakit jantung. Kurang akuratnya peralatan yang digunakan jika hanya mengontrol gula darah dan tekanan darah, dan gaya hidup yang tidak sehat. Data laboratorium yang belum difungsikan secara efektif bisa digunakan untuk deteksi penyakit jantung. Pada penelitian ini akan melakukan deteksi penyakit jantung dengan menggunakan algoritma klasifikasi data mining *naive bayes* berbasis PSO yang akan diukur akurasinya.

II. LANDASAN TEORI

A. Penyakit Jantung

Penyakit jantung merupakan gangguan yang terjadi pada sistem pembuluh darah besar sehingga menyebabkan jantung dan peredaran darah tidak berfungsi sebagaimana mestinya. Penyakit-penyakit yang berhubungan dengan organ jantung dan pembuluh darah antara lain: gagal jantung, jantung koroner, dan jantung rematik [2]. Penyakit jantung koroner (PJK) adalah penyakit jantung dan pembuluh darah yang disebabkan karena penyempitan arteri koroner. Penyempitan pembuluh darah terjadi karena proses *aterosklerosis* atau *spasme* atau kombinasi keduanya. *Aterosklerosis* yang terjadi karena timbunan kolesterol dan jaringan ikat pada dinding pembuluh darah secara perlahan-lahan, hal ini sering ditandai dengan keluhan nyeri pada dada. Pada waktu jantung harus bekerja lebih keras terjadi ketidakseimbangan antara kebutuhan dan asupan oksigen, hal inilah yang menyebabkan nyeri dada. Kalau pembuluh darah tersumbat sama sekali, pemasokan darah ke jantung akan terhenti dan kejadian inilah yang disebut dengan serangan jantung. Penyakit jantung sering disebut “*sudden death*”. Seseorang kemungkinan mengalami serangan jantung karena *iskemia miokard* atau kekurangan oksigen pada otot jantung atau sering disebut dengan nyeri dada.

Beberapa faktor yang bisa menimbulkan penyakit jantung antara lain:

1. bertambahnya usia,
2. gaya hidup
3. stres
4. kurangnya waktu istirahat
5. kurangnya berolah raga
6. merokok
7. obesitas
8. dislipidemia

9. permasalahan dalam diagnosa klinis penyakit jantung

B. Naive Bayes

Bayes merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class* [8]. *Bayes* memiliki akurasi dan kecepatan yang sangat tinggi saat diaplikasi ke dalam *database* dengan data yang besar. Berikut teorema *bayes* :

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} \quad (1)$$

Keterangan :

- X : data dengan *class* yang belum diketahui
- H : hipotesis data x merupakan suatu *class* spesifik
- P(H|X) : probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)
- P(H) : probabilitas hipotesis H (*prior probability*)
- P(X|H) : probabilitas X berdasar kondisi pada hipotesis H
- P(X) : probabilitas dari X

C. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) sering digunakan dalam penelitian, karena PSO memiliki kesamaan sifat dengan *Genetic Algorithm* (GA). Keuntungan dari PSO adalah mudah diterapkan dan ada beberapa parameter untuk menyesuaikan. Sistem PSO diinisiasi oleh sebuah populasi solusi acak dan selanjutnya mencari titik optimum dengan cara meng-*update* tiap hasil pembangkitan. Pendekatan yang digunakan lebih sistematis matematika untuk menemukan solusi.

Particle Swarm Optimization (PSO) dirumuskan oleh Edward dan Kennedy pada tahun

1995. Proses pemikiran di balik algoritma ini terinspirasi dari perilaku sosial hewan, seperti burung yang berkelompok atau sekelompok ikan [9]. Tidak seperti GA, PSO tidak memiliki operator evolusi seperti *crossover* dan mutasi. Baris dalam matriks disebut partikel (sama dengan kromosom GA). Mereka mengandung nilai-nilai variabel dan tidak biner yang dikodekan. Setiap partikel bergerak sekitar dipermukaan partikel dengan kecepatan. Setiap pembaharuan kecepatan dan posisi berdasarkan lokasi terbaik dari lokal dan global:

$$V_{i,m}^{new} = W \cdot V_{i,m}^{old} + C_1 \times R \times (P_{i,m}^{local\ best} - X_{i,m}^{old}) + C_2 \times R \times (P_m^{global\ best} - X_{i,m}^{old}) \quad (2)$$

Menghitung kecepatan baru tiap *particle*:

$$X_{i,m}^{New} = X_{i,m}^{Old} + V_{i,m}^{new} \quad (3)$$

Keterangan:

- n : jumlah partikel dalam kelompok
- d : dimensi
- $V_{i,m}$: kecepatan partikel ke-i pada iterasi ke-i
- w : faktor bobot inersia
- C_1, C_2 : konstanta akselerasi (*learning rate*)
- R : bilangan random (0-1)
- $X_{i,m}$: posisi saat ini dari partikel ke-i pada iterasi ke-i
- $Pbest_i$: posisi terbaik sebelumnya dari partikel ke-i
- $Pgbest$: partikel terbaik diantara semua partikel dalam satu kelompok atau populasi.

Particle Swarm Optimization - Naive Bayes:

Untuk meningkatkan kinerja *naive bayes* maka akan digabungkan algoritma dari *naive bayes* dan PSO. Algoritma *naive bayes* dan PSO diimplementasikan sebagai berikut:

1. menginisialisasi segerombolan $P(t)$ sehingga ketika $t = 0$, lokasi $x_i(t)$ masing-masing partikel $P_i \in P(t)$ dalam ruang ultra adalah acak.
2. setiap posisi partikel $x_i(t)$ mengevaluasi kinerja F
3. bandingkan kinerja masing-masing individu dengan kinerja terbaik yang sejauh ini memiliki, jika $F(x_i(t)) > Pbest_i$ kemudian

$$Pbest_i = F(x_i(t)) \quad (4)$$

$$x_{Pbest_i} = x_i(t) \quad (5)$$

4. bandingkan kinerja setiap partikel dengan *global best particle*, jika $F(x_i(t)) > gbest_i$ kemudian

$$gbest_i = F(x_i(t)) \quad (6)$$

$$x_{gbest_i} = x_i(t) \quad (7)$$

Mengubah vektor kecepatan partikel

$$V_i(t) = V_i(t-1) + \rho_1 \times (x_{Pbest_i} - x_i(t)) + \rho_2(x_{gbest_i} - x_i(t)) \quad (8)$$

Dimana ρ_1 dan ρ_2 adalah variabel acak. ρ_1 dan ρ_2 adalah diidentifikasi sebagai $\rho_1 = r_1 C_1$, $\rho_2 = r_2 C_2$, $r_1, r_2 \sim U(0,1)$, sementara yang C_1 dan C_2 adalah konstanta percepatan positif. lanjutkan ke langkah 2, ulangi rekursi sampai konvergensi. Berikut rumus masukan setiap partikel ke lokasi baru :

$$x_i(t) = x_i(t-1) + v_i(t) \quad (9)$$

$$t = t + 1 \quad (10)$$

5. jika perhitungan bobot dari catatan akhir dalam sampel kumpulan data selesai. Kemudian menghitung bobot dari catatan berikutnya, putar ke langkah 1.
6. hitung rata-rata bobot.

III. HASIL DAN PEMBAHASAN

Penelitian ini merupakan penelitian eksperimen. Pengumpulan data pada penelitian ini meliputi studi literatur berupa buku, jurnal, dan

karya ilmiah yang relevan dengan deteksi penyakit jantung dan data set yang digunakan adalah hasil rekam *medical check-up* yang meliputi hasil laboratorium sejumlah 300 orang yang diolah dengan *tool rapid miner* dan dari data tersebut akan dibagi 75% untuk data *training* dan 25% untuk data *testing* oleh *rapid miner* dengan menggunakan algoritma *naive bayes* berbasis PSO. Data set yang digunakan adalah data hasil laboratorium dan hasil rekam jantung. Untuk data hasil laboratorium parameternya sebagai berikut:

1. Usia
2. Jenis kelamin
3. Hasil darah, meliputi : kolesterol, HDL, LDL, trigliserid, rasio kolesterol
4. Thorax
5. Hasil rekam jantung, meliputi : irama, frekwensi, zona transisi, axis, gelombang P, gelombang QRS, interval P-R, interval QRS, voltage, segmen, gelombang T, gelombang U
6. Cepat capek/nyeri dada
7. Merokok
8. Hipertensi
9. Tekanan darah

Tahapan dalam penelitian:

1. Pada eksperimen awal, dilakukan uji coba parameter dengan menggunakan algoritma *naive bayes* dihasilkan akurasi 82.14%.
2. Eksperimen selanjutnya menggunakan metode *naive bayes* berbasis *particle swarm optimization*. Dari eksperimen dihasilkan akurasi 92.86%.

Dari hasil eksperimen dengan menggunakan *software Rapidminer* dapat diringkas seperti tabel dibawah ini:

Tabel 1. Hasil eksperimen

No	Algoritma	Akurasi	Precision	Recall	AUC
1	<i>Naive bayes</i>	82.14%	23.91%	36%	0.686
2	<i>Naive bayes</i> berbasis PSO	92.86 %	80%	40%	0.839

IV. KESIMPULAN

Pada eksperimen awal dihasilkan akurasi untuk algoritma *naive bayes* sebesar 82.14% dengan nilai *area under cover* (AUC) 0.686 dengan kategori "*poor classification*". Pada eksperimen kedua dengan menggunakan algoritma *naive bayes* berbasis PSO menjadi 92.86% dan nilai AUC 0.839 dengan kategori "*good classification*". Pada eksperimen kedua terbukti bahwa dengan penambahan optimasi dapat meningkatkan nilai akurasi. Penelitian masih perlu dilakukan penelitian dengan menggunakan data yang lebih banyak dan menggunakan metode data mining yang lain.

REFERENSI

- [1] Thomas R Freiden and Donald Berwick, "The "Million Hearts" Initiative - Preventing Heart Attacks," *The New England Journal of Medicine*, September 2010.
- [2] Sutanto, *Cegah dan Tangkal Penyakit Modern Hipertensi, Stroke, Jantung, Kolesterol, dan Diabetes*, Hernita P, Ed. Yogyakarta, Indonesia: Andi Offset, 2010.
- [3] Asha Rajkumar and G.Shophia Reena, "Diagnosa of Heart Disease Using Datamining Algorithm," *Global Journal of Computer Science and Technology*, vol. 10, no. 10, p. 6, September 2010.
- [4] K. Srinivas, G. Raghavendra Rao, and Govardhan, "Survey On Prediction of Heart Morbidity using Data Mining Tehnique ," *International Journal of Data Mining & Knowledge Management Process*, vol. 1, Mei 2011.
- [5] Subbalakshmi, Ramesh, and Chinna Rao, "Decision Support in Heart Disease Prediction System Using Naive Bayes," *Indian Journal of Computer Science and Engineering*, vol. 2, p. 7, April-May 2011.
- [6] Salem. (2010, Agustus) Revealling the silent killer : Coronary Heart Disease. document.
- [7] Kusri and Taufiq Emha Luthfi, *Algoritma Data Mining*, Theresia Ari P., Ed. Yogyakarta, Indonesia: Andi Offset, 2009.
- [8] Budi Santosa, "Tutorial Particle Swarm Optimization," Institut Teknologi Surabaya, Surabaya, 2010.