DOI: 10.24176/detika.v2i1.6964 p-ISSN: 2746-2811

e-ISSN: 2774-2148

OPTIMASI PARTICLE SWARM OPTIMIZATION PADA NAÏVE BAYES UNTUK PREDIKSI KONDISI KELAHIRAN BAYI

Handini Arga Damar Rani¹

¹Universitas IVET Email: ¹hani.arga@gmail.com

(Naskah masuk: 1 Desember 2021, diterima untuk diterbitkan: 8 Desember 2021)

Abstrak

Tehnik data mining dapat digunakan dalam berbagai bidang salah satunya dalam aspek data mining, buat memperkirakan sebuah penyakit dari data rekam medis pasien. Teknik riset yang dipakai pada riset ini mengikuti berbagai tahapan model "Cross-Industry Standard Process Data Mining" (CRISP-DM). Melalui metode klasifikasi dalam data mining, atribut data seperti usia, tekanan darah, berat badan, posisi janin, dan tinggi fundus uteri bisa dipakai buat memperkirakan kemungkinan penyakit pasien. Maka dari itu, peneliti menggunakan metode klasifikasi Naive Bayesian dan optimasi "Particle Swarm Optimization" (PSO) untuk prediksi kelahiran bayi guna mengecek prediksi status kelahiran bayi. Dari hasil prediksi itu bisa dipakai buat menetapkan rata-rata hasil kelahiran bayi setiap bulannya. Data yang peneliti pakai adalah jumlah ibu hamil 165. Selama pengujian digunakan perhitungan akurasi, akurasi, recall, dan AUC chart, dan model prediksi dievaluasi menggunakan 10 "fold cross-validation". Dengan nilai akurasi 91,82% dan precission 100% serta recall 81,50% dan nilai AUC 0.90 termasuk kategori excellent classification pada model yang diujikan.

Kata kunci: prediksi kelahiran bayi, naïve bayes, particle swarm optimization, klasifikasi

OPTIMIZATION OF PARTICLE SWARM OPTIMIZATION IN NAVE BAYES FOR PREDICTION OF BABY BIRTH CONDITIONS

Abstract

Data mining techniques may be utilized in a variety of disciplines, including predicting an illness using patient medical record data, which is one of them. The steps of the "Cross-Industry Standard Process for Data Mining" (CRISP-DM) model were followed in this study's research technique. Data factors like as age, blood pressure, weight, fetal position, uterine fundal height, and other attributes may be utilized to forecast the chance of a patient contracting an illness using the classification process in data mining. As a result, the researcher looked at predicting the baby's birth condition utilizing the naive Bayes classification approach and the "Particle Swarm Optimization" (PSO) optimization, both of which work to forecast a baby's delivery. The average outcomes of infant births each month may be calculated using the predicted results. The prediction model is assessed using 10 fold cross validation in the testing procedure utilizing the computation of accuracy, precision, recall, and AUC diagrams. The data we utilize totals 165 pregnant women. It is included in the good classification category in the model being evaluated, with an accuracy value of 91.82 percent, 100 percent precision, 81.50 percent recall, and an AUC value of 0.90.

Keywords: baby birth prediction, nave bayes, particle swarm optimization, clasification

1. PENDAHULUAN

Data mining ialah proses pemakaian statistik, matematika, kecerdasan buatan, dan teknik *machine leraning* guna mengekstrak dan mengenali informasi yang berguna dan pengetahuan berkorelasi dari bermacam *database* besar (Kusrini, 2019).

Data mining terutama dipakai buat mencari pengetahuan yang ada pada database yang besar,

sehingga biasa disebut Knowledge Discovery in Databases (KDD). Proses penemuan wawasan ini memakai bermacam metode pembelajaran komputer guna (machine learning) menganalisa dan Proses mengekstraknya (Pramudiono, 2006). pencarian bersifat iteratif dan interaktif guna memperoleh konsep atau bentuk yang efektif, baru, berguna dan dapat dipahami. Pada prakteknya, data mining membutuhkan bermacam software analisis

data guna memperoleh konsep dan hubungan data guna melakukan prediksi (Han, 2006).

Masing-masing orang tua harus melakukan yang sebaik mungkin agar kelahiran dapat terlaksana dengan lancar serta bayi bisa lahir dengan sempurna. Prediksi kehamilan berisiko mengacu pada situasi di mana kehamilan akan berdampak negatif pada keselamatan ibu dan bayi. Sejak awal kehamilan, kehamilan berisiko tinggi bisa diperkirakan oleh berbagai faktor seperti keadaan raga serta psikis ibu hamil, asupan gizi, dan penyakit bawaan. Ibu hamil yang termasuk pada kategori ibu hamil risiko tinggi harus waspada dan merencanakan persalinan dini. Sebab banyak ibu hamil yang tak menyadari risiko kehamilan di usia yang sangat muda (Wiknojosastro, 2000).

Menurut Survei Demografi dan Kesehatan Indonesia (SDKI) 2012 angka kematian ibu di Indonesia masih tinggi yaitu 359 per 100.000 kelahiran hidup, dan tujuan global kelima dari MDGs (Millenium Development Goals) adalah menurunkan angka kematian ibu (AKI) yaitu 102 per 100.000 kelahiran hidup (Depkes RI, 2005).

Perkiraan pada kehamilan berisiko dapat diamati dari data rekam medis pasien, sehingga penulis berharap dapat secara mandiri membuat sistem perkiraan status kelahiran bayi lewat sistem aplikasi. Informasi diperoleh dari sistem aplikasi berbasis kebidanan.

Seperti riset yang pernah dilaksanakan oleh Ari Muzakir dan Rika Anisa tahun 2016 yang berjudul "Model data mining sebagai prediksi penyakit hipertensi kehamilan dengan teknik decision tree" Hasil dari riset ini yaitu pohon keputusan dan aturan yang bisa memperkirakan hipertensi pada kehamilan. Supplied Test Set yang disediakan dipakai guna evaluasi memakai WEKA. Hasilnya memiliki tingkat kesalahan sebesar 7,3427% dan tingkat akurasi sebesar 92,6573%. Data training sebanyak 286 instance, yang memperlihatkan terdapat 265 instance akurat dan 21 instance dengan error atau prediksi yang salah (Ari, 2016).

Riset yang dilaksanakan oleh Hilda Amalia dan Evicienna tahun 2017 dengan judul "Komparasi metode data mining untuk penentuan proses persalinan ibu melahirkan" (Amalia, 2017) Dengan membedakan sebagian teknik klasifikasi ialah teknik Naive Bayes, C4.5 dan Neural Network, hasil riset ini menyatakan kalau teknik dengan nilai akurasi terbesar adalah Naive Bayes yakni 94%, dan nilai AUC terbesar pula 0,971. Kinerja AUC dibagi menjadi lima golongan, yaitu: 0,90-1,00 artinya very good classification; 0.80-0.90 artinya good classification; 0.70-0.80 artinya klasifikasi bagus; 0.60-0.70 artinya loose classification (Hilda, 2017). Pada riset itu, nilai AUC Naive Bayes dan neural network tergolong dalam kategori sangat baik, sedangkan algoritma C4.5 tergolong pada kategori bagus.

Pada riset yang sudah dilaksanakan sebelumnya oleh Naisha dengan judul "Aplikasi Prediksi Usia Kelahiran Bayi Dengan Metode Naïve Bayes" Pada tahun 2018, tingkat akurasi tertinggi adalah 78,69%, tingkat precision tertinggi 70,14%, dan tingkat recall tertinggi 63,64% (Indraswari, 2018). Dalam hal penggunaan algoritma naive Bayes untuk menentukan usia lahir bisa memperoleh nilai presisi, recall dan akurasi yang lebih tinggi dibandingkan dengan menggunakan algoritma C.45 untuk menentukan usia lahir (Kurniawan, 2018).

Selain itu dalam riset Desy Kartika I. dengan judul "Aplikasi Prediksi Usia Kelahiran Dengan Metode K-Nearest Neighbor" pada tahun 2019, memakai teknik K-Nearest Neighbor. Variabel yang dipakai yaitu bagian dari faktor pokok yang dialami seorang ibu selama kehamilan, antara lain: usia ibu, tekanan darah, jumlah bayi, riwayat kelahiran, keguguran/kuretase, gizi buruk, penyakit bawaan sebelum kehamilan, dan masalah selama kehamilan. Hasil riset ini disajikan dalam bentuk aplikasi perkiraan usia lahir, dengan tingkat akurasi tertinggi 100% dan tingkat recall tertinggi 84.905660377358% (Indahsari, 2018).

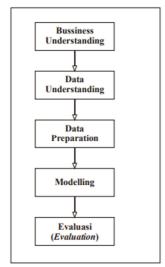
Ada juga riset yang dilaksanakan Jefi dengan judul "Prediksi Bayi Lahir Secara Prematur Dengan Menggunakan Metode C.45 Berbasis Particle Swarm Optimization Pada Klinik Umi", ketika tahun 2019. melalui model algoritma C4.5 dan model algoritma C4.5 berbasis particle swarm optimization, diperoleh aturan prediksi kelahiran prematur untuk menyajikan nilai akurasi yang lebih akurat. Sesudah dilaksanakan pengetesan kedua model algoritma C4.5 dan algoritma C4.5 berbasis particle swarm optimization, didapatkan hasil akurasi algoritma C4.5 sebesar 94,30%, dan nilai AUC sebesar 0,986 dengan tingkat akurasi yang sangat baik. klasifikasi dan diagnosis, tetapi Setelah satu, algoritma C4.5 berdasarkan optimasi segerombolan partikel memiliki tingkat akurasi 97,91% dan nilai AUC 0,997, yang memiliki tingkat klasifikasi dan diagnosis yang sangat baik. Alhasil kedua teknik itu mempunyai perbedaan jenjang akurasi ialah sekitar 3.61% (Jefi, 2019).

Selain itu penelitian yang dilakukan oleh (Handini, 2020) yang berjudul "Sistem Prediksi Kondisi Kelahiran Bayi Menggunakan Klasifikasi Naïve Bayes" dengan hasil akurasi 80,91% dan nilai recall 85,50% dan precision 75,45%, serta nilai AUC 0.91.

2. METODE PENELITIAN

Teknik riset yang diterapkan dalam riset ini melalui langkah-langkah bentuk "Cross-Industry Standard Process for Data Mining" (CRISP-DM). Terdapat juga langkah-langkah CRISP-DM, yaitu "Busssiness Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment" (CRISP, 2009). Pada riset ini Cuma

membutuhkan langkah-langkah seperti pada gambar 1 di bawah ini:



Gambar 1. Diagram Alur Penelitian

2.1 Bussiness Understanding

Pemahaman bisnis ialah memahami tujuan bisnis, mengevaluasi keadaan serta mengubah tujuan bisnis menjadi tujuan penambangan data. Dalam penelitian ini membutuhkan pengetahuan untuk mendapatkan data untuk memprediksi kelahiran ibu hamil apakah beresiko atau normal, dengan memakai database kelahiran yang berlangsung pada periode 1 tahun terakhir dalam suatu klinik bersalin.

2.2 Data Understanding

Pada langkah ini dilaksanakan proses penghimpunan data, lalu data tersebut dianalisis dan kualitas data yang dipakai pada riset ini dievaluasi. Sumber data yang dipakai pada riset ini ialah informasi yang ada pada suatu klinik bersalin di kota Kudus, pada Januari 2020- data 2021, Riset ini pula memakai data dari data primer bisa diamati dalam tabel 1 di bawah ini:

Tabel 1. Contoh data set kelahiran									
No_	Nama	Nama	ala	T	В	Us	Letak	TFU	Label
RM		Suami	ma	D	В	ia	janin		
			t				-		
	Mu'alif	Hadi	Ku	1	45	20	Sesuai	Sesuai	Normal
P 1	atun		du	0					
			s	0					
	Yunita	Efendi	Ku	1	56	22	Sesuai	Sesuai	Normal
P_2			du	2					
			s	0					
	Elis	Yanto	Ku	1	50	21	Sesuai	Tidak	Resiko
P_3			du	0				sesuai	
			s	0					
	Miftakh	Deni	Ku	1	55	25	Sesuai	Sesuai	Normal
P 4	ul		du	1					
			s	0					
	Suyanti	Agus	Ku	1	59	32	Sesuai	Sesuai	Normal
P_5			du	1					
			s	0					

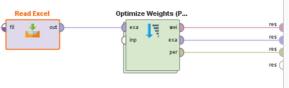
2.3 Data Preparation

Pada langkah ini mencakup pemilihan data dan pemilihan atribut-atribut atau field yang sudah terpilih yang akan digunakan nantinya untuk proses klasifikasi, menjadi 1 tabel seperti pada tabel 2 di bawah ini:

Tabel 2. Data dengan atribut yang dipilih						
Nama	TD	BB	Usia	Letak janin	TFU	label
				Janni		
Mu'alif atun	100	45	20	Sesuai	Sesuai	Normal
Yunita	120	56	22	Sesuai	Sesuai	Normal
Elis	100	50	21	Sesuai	Tidak sesuai	Resiko
Miftak hul	110	55	25	Sesuai	Sesuai	Normal
Suyanti	110	59	32	Sesuai	Sesuai	Normal

2.4 Modelling

Pada tahap ini, bermacam teknik pemodelan dipilih dan diaplikasikan pada kumpulan data yang disiapkan guna memenuhi keperluan bisnis yang sesuai. Teknik yang dipakai adalah teknik klasifikasi, melalui pemakaian algoritma Naïve bayes dan Optimasi Particle Swarm Optimization (PSO). Tahap pemodelan pula meliputi evaluasi dan analisis komparatif dari bermacam model yang dibentuk menggunakan software Rapid Miner dapat dilihat pada gambar 2 sebagai berikut:



Gambar 2. Model analisis data menggunakan Rapid Miner

Naïve bayes

Bayes ialah klasifikasi statistik yang bisa dipakai guna memperkirakan probabilitas anggota suatu kelas (Budi, 2010). Bayes memiliki akurasi dan kecepatan yang amat besar ketika diaplikasikan dalam database big data. Ini adalah teorema Bayes:

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

Keterangan:

= Data dengan class yang belum diketahui \mathbf{x}

= Hipotesis data x merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (posteriori probability)

= Probabilitas hipotesis H (prior probability) P(H)

P(X|H) = Probabilitas X berdasar koondisi pada hipotesis H

P(X)= Probabilitas dari X

Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) sering digunakan dalam penelitian karena Particle Swarm

Optimization (PSO) memiliki karakteristik yang mirip dengan Genetic Algoritm (GA). Kelebihan PSO adalah mudah diimplementasikan dan memiliki beberapa parameter yang dapat disesuaikan. Sistem PSO dimulai dengan serangkaian solusi acak, dan kemudian menemukan titik terbaik dengan memperbarui setiap hasil yang dihasilkan. Metode vang digunakan adalah matematika vang lebih sistematis untuk mencari solusi. Particle Swarm Optimization (PSO) dikembangkan oleh Edward serta Kennedy pada tahun 1995. Cara pandangan di balik algoritma ini termotivasi oleh sikap sosial binatang, semacam kawanan burung ataupun kawanan ikan (Budi, 2010). Berbeda dengan GA, PSO tidak mempunyai operator evolusioner semacam crossover serta pemindahan. Baris dalam dituturkan partikel (serupa matriks kromosom GA). Mereka memiliki nilai- nilai variabel serta bukan kode biner. Tiap partikel beranjak di sekitar permukaan partikel dengan kecepatan. Kecepatan dan lokasi masing-masing pembaruan didasarkan pada lokasi lokal dan global terbaik:

$$\begin{split} V_{i,m}^{new} &= W.V_{i,m}^{old} + C_1 \, x \left(P_{i,m}^{local \ best} - X_{i,m}^{old} \right) + \\ C_2 \, x \, R \, x \left(P_{i,m}^{global \ best} - X_{i,m}^{old} \right) \end{split} \tag{2}$$

Mengukur kecepatan terkini masing-masing particle:

$$X_{i,m}^{new} = X_{i,m}^{old} - V_{i,m}^{new}$$
 (3)

Keterangan:

n: jumlah partikel dalam kelompok

d: dimensi

V, : kecepatan partikel ke-i pada iterasi ke-i

w: faktor bobot inersia

C1, C2: konstanta akselerasi (learning rate)

R: bilangan random (0-1)

X.:posisi saat ini dari partikel ke-i pada iterasi ke-i Pbesti: posisi terbaik sebelumnya dari partikel ke-i Pgbest: partikel terbaik diantara semua partikel dalam satu kelompok atau populasi.

2.5 Evaluasi

Dalam tahap ini, model yang telah dirancang , diuji dan dievaluasi keakuratan dan generalitasnya. Menggunakan tehnik perhitungan nilai akurasi, recall, precision, AUC dan validasi menggunakan 10 fold cross validation. Langkah ini menghitung sejauh mana model yang dipilih memenuhi tujuan bisnis, dan jika demikian, sejauh manakah itu (apakah lebih banyak model perlu dibuat dan diukur).

3. HASIL DAN PEMBAHASAN

Riset ini ialah riset eksperimen. Pengumpulan data dalam riset ini mencakup pengamatan serta tanya jawab di klinik kebidanan dan kandungan di Kudus. Dari hasil pengamatan dan tanya jawab didapat data berbentuk data catatan ibu hamil yang berjumlah 165 ibu hamil. Data yang digunakan adalah data pasien hamil selama 1 tahun. Setelah itu, dimasukkan data pada format excel. Data dalam database seringkali tidak semua atribut data yang digunakan, sehingga hanya atribut data yang sesuai guna dianalisa yang bakal diperoleh dari database, setelah itu diklasifikasikan memakai teknik naive bayes, dan dioptimasi memakai teknik Particle Swarm Optimization (PSO), yang dikelola dengan tool k Miner, 75% data bakal disajikan menjadi data training, dan 25% bakal dibagi menjadi data testing oleh Rapid Miner memakai algoritma naive Bayes berbasis PSO.

Database yang dipakai ialah data hasil pemeriksaan dokter. Untuk data hasil pemeriksaan kehamilan parameternya yang digunakan sebagai berikut:

- 1. Nama
- 2. Tekanan Darah
- 3. Berat Badan
- 4. Usia
- 5. Letak Janin
- 6. Tinggi Fundus Uterus (TFU)
- 7. Label

Selanjutnya Eksperimen memakai teknik naive bayes berbasis particle swarm optimization. Dalam eksperimen tersebut diperoleh akurasi sebesar 91,82%. Dalam hasil eksperimen dengan memakai software Rapid Miner bisa disimpulkan seperti table 3 di bawah ini:

Tabel 3. Hasil Akurasi menggunakan confusion matrik

Accuracy: 91,82% +/- 7,96%(micro average: 91,82%)

	True Normal	True Resiko	Class	
			Precission	
Prediksi	62	9	97 22 6/	
Normal	02	9	87,32%	
Prediksi	0	39	1000	
Resiko	0	39	100%	
Class Recall	100%	81,25%	-	

Untuk nilai class precision yaitu: 100%, dapat dilihat dalam tabel 4 berikut:

Tabel 4. Hasil class precision

Precision: 100% +/- 0.00%

	True Normal	True Resiko	Class	
			Precission	
Prediksi	62	9	87,32%	
Normal	02	9	81,32%	
Prediksi	0	39	100%	
Resiko	U	39	100%	
Class Recall	100%	81,25%	-	

Untuk nilai class recall yaitu: 81,50%, dapat dilihat dalam tabel 5 berikut:

Tabel 5. Hasil class recall

recall: 81.50% +/- 17.65%

	True Normal	True Resiko	Class	
			Precission	
Prediksi	62	0	87,32%	
Normal	02	9	67,32%	
Prediksi	0	39	100%	
Resiko	U	39	100 /6	
Class Recall	100%	81,25%	-	

Nilai Auc (Area Under Curve) ialah sekitar 0.906. Berlandaskan gambar itu, teknik yang dipakai pada riset ini tercantum pada kategori Excellent Classification. Bisa diamati pada gambar 3 berikut:



Gambar 3. Nilai Area Under Curve (AUC)

4. KESIMPULAN

Berlandaskan hasil riset dari langkah-langkah pemahaman bisnis, pemahaman data, penyiapan data, pemodelan dan evaluasi, alhasil ada sebagian hal yang diambil kesimpulan yaitu:

- 1. Penggunaan algoritma klasifikasi naïve bayes berbasis PSO pada eksperimen menghasilkan nilai akurasi 91.82% dan nilai AUC 0.906 dengan kategori "excellent classification".
- 2. Dengan memakai hasil pemeriksaan dokter, sistem perkiraan mampu menentukan rata-rata persalinan normal atau persalinan berbahaya per
- 3. Riset pula harus dilaksanakan dengan memakai lebih banyak data dan metode data mining lainnya.

REFERENCES

- KUSRINI DAN EMHA TAUFIQ LUTHFI. 2009. Algoritma Data Mining. Penerbit Andi. Yogyakarta.
- PRAMUDIONO, I. 2006. Apa itu Data Mining? http://datamining.japati.net/

- cgibin/indodm.cgi. Diakses tanggal 26 November 2015.
- HAN J DAN KAMBER M. 2006. Data mining Concepts and Techniques 2nd Edition. The Morgan Publisher. Kaufmann Fransisco.
- H. WIKNJOSASTRO, 2000. "Ilmu Kebidanan", Jakarta:PT. Pustaka Bina Sarwono Prawirohardjo,
- Tim Kajian AK. 2005."Kajian kematian ibu dan anakdi Indonesia," Badan Penelitian danPengembangan Kesehatan. Jakarta: Depkes RI,.
- B. MAX, 2007, Principles of Data Mining, Undergraduate Topics in Computer Science ISSN: 1863-7310, London: Springer,
- ANASTASIA HASNI G. DAN SETIADI TEDY, "Penerapan data mining untuk mengetahui factor-faktor yang mempengaruhi kelahiran bayi menggunakan association rules", Jurnal Sarjana Teknik Informatika e-ISSN: 2338-5197 Volume 2 Nomor 3, Oktober 2014.
- MUZAKIR ARI DAN WULANDARI RIKA ANISA, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree", Scientific Journal of Informatics , Vol. 3, No. 1, Mei 2016.
- AMALIA HILDA DAN EVICIENNA, "Komparasi Metode Data Mining Untuk Penentuan Proses Persalinan Ibu Melahirkan", Jurnal Sistem Informasi (Journal of Information System), Volume 13, Issue 2, October 2017
- F. GORUNESCU, "Data mining Concept and Technique", Verlag Berlin Heidelberg: Springer, 2011.
- INDRASWARI, N. R., & KURNIAWAN, Y. I. (2018). Aplikasi Prediksi Usia Kelahiran Dengan Metode Naïve Bayes. Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, 9(1), 129-138.
- KURNIAWAN, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C. 45 dalam Data Mining. Klasifikasi Jurnal TeknologiInformasi dan Ilmu Komputer, 5(4), 455-464.
- INDAHSARI DESY KARTIKA, KURNIAWAN YOGIEK INDRA, 2019, Aplikasi Prediksi Usia Kelahiran Dengan Metode K-Nearest Neighbor, Jurnal Kebidanan, Vol. XI, No. 01, Juni 2019.
- JEFI, 2019, Prediksi Bayi Lahir Secara Prematur Dengan Menggunakan Metode C.45 Berbasis Particle Swarm Optimization Pada Klinik Umi, Indonesian Journal on

- Networking and Security Volume 8 No 2 -2019.
- ALFIRONI BURHAN, 2013, Implementasi Data Mining Dengan Naive Bayes Classifier Untuk Mendukung Strategi Pemasaran Di bagian Humas Stmik Amikom Yogyakarta, Yogyakarta.
- HANDINI ARGA D, 2020, "Sistem Prediksi Kondisi Kelahiran Bayi Menggunakan Klasifikasi Naïve Bayes", Joined Journal Vol 3, No. 2. Desember 2020
- BUDI SANTOSA, 2010, "Tutorial Particle Swarm Optimization," Institut Teknologi Surabaya, Surabaya.