

Seleksi Fitur Klasifikasi Penyakit Diabetes Menggunakan Particle Swarm Optimization (PSO) Pada Algoritma Naive Bayes

Nurlaelatul Maulidah¹, Ari Abdilah², Elah Nurlalah³, Windu Gata⁴, Fuad Nur Hasan⁵

^{1,2,4} Jurusan Ilmu Komputer, STIMIK Nusa Mandiri

Jalan Margonda Raya No. 545, Pondok Cina, Depok, Jawa Barat 16424, (021) 31908575

e-mail: 14002377@nusamandiri.ac.id¹, ari.aab@nusamandiri.ac.id², windu@nusamandiri.ac.id³

³Jurusan Sistem Informasi, Universitas Bina Sarana Informatika

Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450, (021) 21231170, e-mail: elah.enl@bsi.ac.id

⁵Jurusan Ilmu Komputer, Universitas Bina Sarana Informatika

Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450, (021) 21231170, e-mail: fuad.fnu@bsi.ac.id

ARTICLE INFO

Article history:

Received 21 November 2020

Received in revised form 22 November 2020

Accepted 27 November 2020

Available online 10 Desember 2020

ABSTRACT

Diabetes is a serious chronic disease that occurs because the pancreas does not produce enough insulin (a hormone that regulates blood sugar or glucose), or when the body cannot effectively use the insulin it produces. WHO data shows that the incidence of non-communicable diseases in 2004 reached 48 , 30% is slightly higher than the incidence rate of infectious diseases, namely 47.50% [1]. According to the Ministry of Health in 2012 diabetes caused 1.5 million deaths. Some Indonesian people, this disease is better known as diabetes or blood sugar. This research was developed through secondary data processing from the Pima Indians Diabetes Dataset health database which was taken from the Kaggle dataset and can be accessed through <https://www.kaggle.com/uciml/pima-indians-diabetes-database>. Where the data itself consists of 768 records with several medical predictor variables (Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age and Outcome). Then the data will be processed using the Particle Swarm Optimization (PSO) feature selection to increase the accuracy value and the Naive Bayes algorithm to determine the accuracy results of the diagnosis of diabetes. From the results of research that has been done for the accuracy of the classification algorithm Naive Bayes is 74.61%, while the accuracy of the classification algorithm with Particle Swarm Optimization is 77.34% with an accuracy difference

of 2.73%. So it can be concluded that the application accuracy than using only the individual method, of the Particle Swarm Optimization technique is able namely the Naive Bayes algorithm. to select attributes in the Naive Bayes Algorithm, **Keywords:** Diabetes, Particle Swarm Optimization, and can produce a better level of diabetes diagnosis Naive Bayes Algorithm

ABSTRAK

Diabetes adalah penyakit kronis serius yang terjadi karena pankreas tidak menghasilkan cukup insulin (hormon yang mengatur gula darah atau glukosa), atau ketika tubuh tidak dapat secara efektif menggunakan insulin yang dihasilkannya. Data WHO menunjukkan bahwa angka kejadian penyakit tidak menular pada tahun 2004 yang mencapai 48,30% sedikit lebih besar dari angka kejadian penyakit menular, yaitu sebesar 47,50% [1]. Menurut kementerian kesehatan pada tahun 2012 penyakit diabetes menyebabkan 1,5 juta kematian. Sebagian masyarakat Indonesia, penyakit ini lebih dikenal dengan nama kencing manis atau gula darah. Penelitian ini dikembangkan melalui pengolahan data sekunder dari *database* kesehatan *Pima Indians Diabetes Dataset* yang diambil dari dataset *Kaggle* dan dapat diakses melalui <https://www.kaggle.com/uciml/pima-indians-diabetes-database>. Dimana datanya sendiri terdiri dari 768 record dengan beberapa variabel prediktor medis (*Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age dan Outcome*). Kemudian data tersebut nantinya diolah menggunakan seleksi fitur *Particle Swarm Optimization* (PSO) untuk meningkatkan nilai akurasi dan algoritma *Naive Bayes* untuk mengetahui hasil akurasi dari diagnosis penyakit diabetes. Dari hasil penelitian yang telah dilakukan untuk akurasi algoritma klasifikasi *Naive Bayes* sebesar 74,61%, sedangkan untuk akurasi algoritma klasifikasi dengan *Particle Swarm Optimization* sebesar 77,34% dengan selisih akurasi sebesar 2,73%. Sehingga dapat disimpulkan bahwa penerapan teknik optimasi *Particle Swarm Optimization* mampu menyeleksi atribut pada Algoritma *Naive Bayes*, dan dapat menghasilkan tingkat akurasi diagnosis penyakit diabetes yang lebih baik dibandingkan hanya menggunakan metode individual yaitu algoritma *Naive Bayes*.

Kata Kunci: Diabetes, *Particle Swarm Optimization*, Algoritma *Naive Bayes*

1. PENDAHULUAN

Diabetes adalah penyakit kronis serius yang terjadi karena pankreas tidak menghasilkan cukup insulin (hormon yang mengatur gula darah atau glukosa), atau ketika tubuh tidak dapat secara efektif menggunakan insulin yang dihasilkannya. Data WHO menunjukkan bahwa angka kejadian penyakit tidak menular pada tahun 2004 yang mencapai 48,30% sedikit lebih besar dari angka kejadian penyakit menular, yaitu sebesar 47,50% [1].

Toharin, et. Al mengemukakan bahwa Diabetes sering disebut dengan *mother of disease* karena dapat menjadi penyebab munculnya penyakit komplikasi seperti hipertensi, penyakit jantung dan pembuluh darah, stroke, gagal ginjal, dan kebutaan, serta kerusakan jangka panjang meliputi gangguan dan kegagalan fungsi berbagai organ terutama mata, ginjal, syaraf, jantung, dan pembuluh darah [2].

Secara global, diperkirakan 422 juta orang dewasa hidup dengan diabetes pada tahun 2014, dibandingkan dengan 108 juta pada tahun 1980. Prevalensi diabetes di dunia (dengan usia yang distandarisasi) telah meningkat hampir dua kali lipat sejak tahun 1980, meningkat dari 4,7% menjadi 8,5% pada populasi orang dewasa. Hal ini mencerminkan peningkatan faktor risiko terkait seperti kelebihan berat badan atau obesitas. Selama beberapa dekade terakhir, prevalensi diabetes meningkat lebih cepat di negara berpenghasilan rendah dan menengah daripada di negara berpenghasilan tinggi. Diabetes menyebabkan 1,5 juta kematian pada tahun 2012. Gula darah yang lebih tinggi dari batas maksimum mengakibatkan tambahan 2,2 juta kematian, dengan meningkatkan risiko penyakit kardiovaskular dan lainnya. Empat puluh tiga persen (43%) dari 3,7 juta kematian ini terjadi sebelum usia 70 tahun. Persentase kematian yang disebabkan oleh diabetes yang terjadi sebelum usia 70 tahun lebih tinggi di negara-negara berpenghasilan rendah dan menengah daripada di negara-negara berpenghasilan tinggi [1].

2. TINJAUAN PUSTAKA

Penelitian ini dikembangkan melalui pengolahan data sekunder dari *database* kesehatan *Pima Indians Diabetes Dataset* yang diambil dari dataset *Kaggle* dan dapat diakses melalui

Seleksi Fitur Klasifikasi Penyakit Diabetes Menggunakan Particle Swarm Optimization (PSO) Pada Algoritma Naive Bayes (Nurlaelatul Maulidah)

<https://www.kaggle.com/uciml/pima-indians-diabetes-database>. Dimana datanya sendiri terdiri dari 768 record dengan beberapa variabel prediktor medis (*Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age dan Outcome*). Kemudian data tersebut nantinya diolah menggunakan seleksi fitur *Particle Swarm Optimization* (PSO) untuk meningkatkan nilai akurasi dan algoritma *Naive Bayes* untuk mengetahui hasil akurasi dari diagnosis penyakit diabetes.

2.1. Particle Swarm Optimization(PSO)

Handika, et al mengemukakan bahwa Algoritma PSO pertama kali diusulkan oleh J. Kennedy dan R. C Eberhart (Kennedy, 1995). Particle Swarm Optimization (PSO) merupakan sebuah metode optimasi yang didasarkan pada perilaku sebuah kawanan serangga misalnya semut, rayap, lebah atau burung. Suatu partikel dalam ruang memiliki posisi dan setiap posisi dalam ruang pencarian merupakan alternatif solusi yang dapat dievaluasi menggunakan fungsi objektif. Setiap partikel dapat menyesuaikan posisi dan kecepatan masing-masing dengan cara setiap partikel menyampaikan informasi terbaiknya kepada partikel yang lain. Oleh karena itu, setiap partikel memiliki kecenderungan untuk terbang menuju posisi yang dianggap terbaik Partikel sendiri adalah setiap individu dalam suatu kelompok [3].

Berikut langkah-langkah untuk melakukan optimasi dengan seleksi fitur *Particle Swarm Optimization* yaitu:

1. Pengumpulan Data

Peneliti menggunakan data yang didapatkan dari database kesehatan *Pima Indians Diabetes Dataset* (PPID) dan hasil pemeriksaan terhadap 768 orang dengan 8 atribut dan satu atribut *class*.

2. Pengolahan Awal Data

Dalam tahap pengolahan awal data diawali dengan proses input data ke format yang dibutuhkan, setelah itu pembersihan data, dilanjutkan dengan pengelompokan data, serta pemecahan data untuk digunakan dalam proses pembelajaran dan pengujian. Dari proses pengumpulan dan pengolahan data awal diperoleh sebanyak 768 data yang terdiri dari 268 positif diabetes dan 500 negatif diabetes.

3. Metode yang diusulkan

Metode yang diusulkan dalam penelitian ini adalah metode klasifikasi data mining yaitu Algoritma *Naive Bayes* dan Algoritma *Naive Bayes* dioptimasi dengan *Particle Swarm Optimization*.

4. Experimen dan pengujian model

Pada penelitian ini Experimen dan pengujian model menggunakan *Cross Validation* yaitu dengan cara dataset dibagi menjadi 10 bagian kemudian dilakukan pengulangan pengujian sebanyak 10 kali pengulangan.

5. Evaluasi dan validasi hasil

Pada penelitian ini evaluasi dan validasi hasil dilakukan terhadap model-model untuk mendapatkan informasi model yang akurat dengan dilakukan pengujian dan menggunakan metode *Confusion Matrix* dan *Curva ROC*.

2.2. Algoritma Naive Bayes

Algoritma Naive Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya[4].

Menurut Bustami dalam Saleh (2015:209) persamaan dari teorema *Bayes* adalah:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots \dots \dots (2.1)$$

Dimana:

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* yang spesifik

$P(H|X)$: Probabilistik hipotesis H berdasar kondisi X (posteriori probabilistik)

$P(H)$: Probabilistik hipotesis H (prior probabilitas)

$P(X|H)$: Probabilistik hipotesis X berdasar kondisi pada hipotesis H

$P(X)$: Probabilitas X

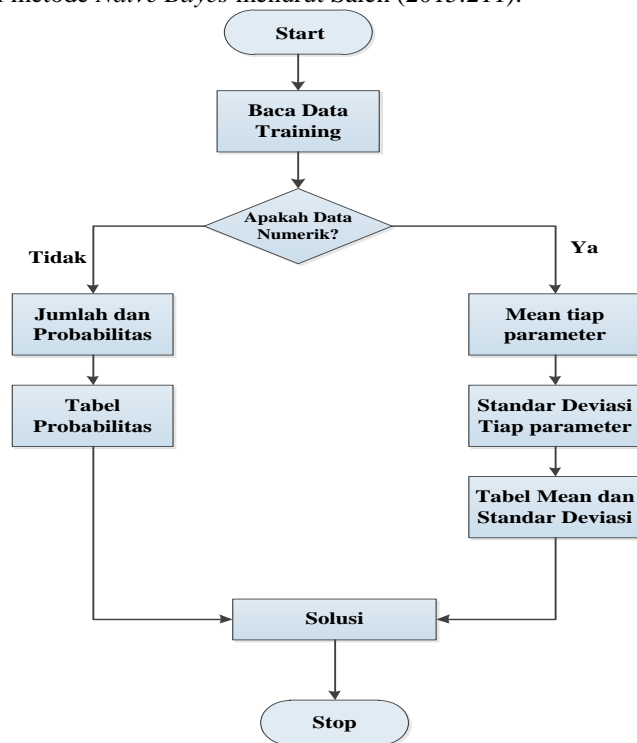
Untuk menjelaskan metode *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naïve Bayes* di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1|C) \dots P(F_n|C)}{P(F_1 \dots F_n)} \dots \dots \dots (2.2)$$

Dimana variable C merepresentasikan kelas, sementara variable $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sample berkarakteristik tertentu kedalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuk sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu rumus diatas dapat ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{prior \times likelihood}{evidence} \dots \dots \dots (2.3)$$

Berikut adalah alur dari metode *Naïve Bayes* menurut Saleh (2015:211):



Sumber: Saleh (2015:211)

Gambar 1. Alur Metode Naïve Baiyes

Adapun keterangan dari gambar diatas sebagai berikut:

1. Baca data training
2. Hitung jumlah dan probabilitas, namun apabila data numerik maka:
 - a. Cari nilai *mean* dan standar deviasi dari masing-masing parameter yang merupakan data numerik. Adapun persamaan yang digunakan untuk menghitung nilai rata-rata hitung (*mean*) dapat dilihat sebagai berikut:

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \dots \dots \dots (2.4)$$

Atau

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \dots \dots \dots (2.5)$$

di mana:

μ : rata-rata hitung (*mean*)

x_i : nilai sample ke i

n : jumlah sampel

Dan persamaann untuk menghitung simpangan baku (standar deviasi) dapat dilihat sebagai berikut:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \dots\dots\dots (2. 6)$$

Dimana:

σ : standar deviasi

x_i : nilai x ke -i

μ : rata-rata hitung

n : jumlah sampel

- b. Cari nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Mendapatkan nilai dari tabel *mean*, standar deviasi dan probabilitas.
4. Solusi kemudian dihasilkan.

3. HASIL

3.1. Hasil Pemodelan Dengan Algoritma Naive Bayes

Untuk dapat membentuk model diperlukan data training, dan tahap modelling menggunakan dengan software rapidminer agar mampu membaca data dengan banyak format, yang paling umum digunakan diantaranya .csv, .xls dan .mdb. Dataset diambil dari Kaggle dalam bentuk format .txt kemudian ditransformasi menjadi format .xls.

3.1.1. Evaluasi Model Dengan Confusion Matrix

Model confusion matrix akan membentuk matrix yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif, kemudian masukan data testing yang sudah disiapkan ke dalam confusion matrix sehingga didapatkan hasil pada tabel di bawah ini:

Tabel 1. Confusion Matrix Algoritma Klasifikasi Naive Bayes

	True YES	True NO	Class Precision
Pred. YES	162	89	64.54%
Pred. NO	106	411	79.50%
Class recall	60,45%	82.20%	

Sumber: (Maulidah et al, 2020)

Berdasarkan tabel diatas dari data testing terdapat rincian jumlah True Positive (TP) 162, False Negative (FN) 89, False Positive (FP) adalah 89 dan True Negative (TN) 411. Dari data tersebut maka dapat dihitung nilai accuracy, sensitvity, specifity, PPV dan NPV.

$$Accuracy = \frac{162+411}{162+411+106+89} = \frac{573}{768} = 0,7461 / 74.61\%$$

$$Sensitivity = \frac{162}{162+89} = \frac{162}{251} = 0,6454 / 64,54\%$$

$$Specificity = \frac{411}{411+106} = \frac{411}{517} = 0,7950 / 79,50\%$$

$$PPV = \frac{162}{162+106} = \frac{162}{268} = 0,6045 / 60,45\%$$

$$NPV = \frac{411}{411+89} = \frac{411}{500} = 0,822 / 82,20\%$$

3.2. Hasil Pemodelan dengan Algoritma Naive Bayes - Particle Swarm Optimization

Pada Particle Swarm Optimization, Perhitungan pada keseluruhan data dilakukan dengan Rapid miner, model yang terbentuk dengan algoritma Naive Bayes sendiri akan dibandingkan dengan model yang terbentuk setelah proses pembobotan atribut. Nilai bobot atribut akan diberikan kepada setiap partikel Particle Swarm Optimization. Model yang terbentuk akan diuji sehingga dapat diketahui akurasi model yang terbentuk, setelah itu partikel Particle Swarm Optimization akan bergerak ke arah partikel terbaik agar akurasi model yang dihasilkan dapat lebih tinggi dari sebelumnya. Atribut yang terseleksi dari 8 atribut menjadi 4 atribut, yang terdiri dari Pregnancies, Blood Pressure, Skin Thickness, dan BMI. Berikut model confusion matrix Algoritma Naive Bayes – PSO pada tabel dibawah ini:

Tabel 2. Confusion Matrix Algoritma Klasifikasi Naive Bayes – PSO

	True YES	True NO	Class Precision
Pred. YES	158	64	71.17%
Pred. NO	110	436	79.85%
Class recall	58,96%	87.20%	

Sumber: (Maulidah et al, 2020)

Berdasarkan tabel diatas dari data testing terdapat rincian jumlah True Positive (TP) 158, False Negative (FN) 64, False Positive (FP) adalah 110 dan True Negative (TN) 436. Dari data tersebut maka dapat dihitung nilai accuracy, sensitivity, specificity, PPV dan NPV.

$$Accuracy = \frac{158+436}{158+436+110+64} = \frac{594}{768} = 0,7734 / 77,34\%$$

$$Sensitivity = \frac{158}{158+64} = \frac{158}{222} = 0,7117 / 71,17\%$$

$$Specificity = \frac{436}{436+110} = \frac{436}{546} = 0,7985 / 79,85\%$$

$$PPV = \frac{158}{158+110} = \frac{158}{268} = 0,5896 / 58,96\%$$

$$NPV = \frac{436}{436+64} = \frac{436}{500} = 0,872 / 87,20\%$$

Maka dapat disimpulkan bahwa algoritma Naive Bayes + Particle Swarm Optimization dapat memprediksi penyakit diabetes Positif dan Negatif dengan akurasi yang baik sebesar 77.34%.

3.3. Tabel Probabilitas

3.3.1. Tabel Probabilitas Model Naive Bayes

Dalam membuat model Naive Bayes terlebih dahulu kita mencari probabilitas hipotesis untuk masing-masing Kelas P(H). Hipotesis yang ada yaitu pasien yang menderita penyakit liver dan pasien yang tidak menderita penyakit liver, perhitungan probabilitas yaitu seperti dibawah ini:

$$P(Yes) = 268 : 768 = 0,3$$

$$P(No) = 500 : 768 = 0,6$$

Setelah probabilitas untuk tiap hipotesis diketahui, langkah selanjutnya adalah menghitung probabilitas kondisi tertentu (probabilitas X) berdasarkan probabilitas tiap hipotesis (probabilitas H) atau dinamakan probabilitas prior.

Tabel 3. Hasil perhitungan probabilitas priordengan menggunakan Naive Bayes.

Atribut	Jml Kasus	Ya	Tidak	P(X Ci)	
		(Si)		Ya	Tidak

		(S)		(Si)		
Total		768	268	500		
Pregnancies	0-5	549	157	392	0,585	0,784
	6-10	185	91	94	0,339	0,188
	11-17	34	20	14	0,074	0,028
Glucose	0-79	41	3	38	0,011	0,076
	80-89	63	4	59	0,014	0,118
	90-99	93	9	84	0,033	0,168
	100-145	407	134	273	0,5	0,546
	>=146	164	118	46	0,440	0,092
BloodPressure	0-80	603	191	412	0,712	0,824
	82-98	149	68	81	0,253	0,162
	100-122	16	9	7	0,033	0,014
SkinThickness	0-20	351	105	246	0,391	0,492
	21-30	179	55	124	0,205	0,248
	31-45	207	92	115	0,343	0,23
	46-99	31	16	15	0,059	0,03
Insulin	0-59	434	143	291	0,533	0,582
	60-79	45	4	41	0,014	0,082
	81-99	39	9	30	0,033	0,06
	>=100	250	112	138	0,417	0,276
BMI	0-23.09	80	7	73	0,026	0,146
	24-28.05	139	28	111	0,104	0,222
	28.06-32.05	137	51	86	0,190	0,172
	32.06-35.09	143	62	81	0,213	0,162
	36.01-39.05	94	39	55	0,145	0,11
	39.06-55	177	83	94	0,309	0,188
Diabetes Pedigree Function	0.0-0.193	321	84	237	0,328	0,474
	0.2-0.399	237	83	154	0,309	0,308
	0.4-0.692	159	70	89	0,261	0,178
	1.0-1.1	10	4	6	0,014	0,012
	1.1-2.329	37	23	14	0,085	0,028
Age	21-40	574	166	408	0,619	0,816
	41-59	162	93	69	0,347	0,138
	60-81	32	9	23	0,033	0,046

Sumber: (Maulidah et al, 2020)

3.3.2. Tabel Probabilitas Model Naive Bayes – PSO

Dalam mengetahui probabilitas dari model naive bayes dengan Particle Swarm Optimization, proses yang dilakukan sama dengan proses pembuatan model sebelumnya yaitu model naive bayes namun perbedaannya adalah dari atribut yang digunakan, atribut yang digunakan adalah atribut yang terpilih dari hasil seleksi atribut terbaik menggunakan Particle Swarm Optimization, berikut hasil probabilitas prior untuk model naive bayes + particle swarm optimization:

Tabel 4. Probabilitas Prior setelah di seleksi menggunakan particle swarm optimization

Atribut	Jml Kasus (S)	Ya (Si)	Tidak (Si)	P(X Ci)		
				Ya	Tidak	
Total	768	268	500			
Pregnancies	0-5	549	157	392	0,585	0,784
	6-10	185	91	94	0,339	0,188
	11-17	34	20	14	0,074	0,028
BloodPressure	0-80	603	191	412	0,712	0,824
	82-98	149	68	81	0,253	0,162
	100-122	16	9	7	0,033	0,014
SkinThickness	0-20	351	105	246	0,391	0,492
	21-30	179	55	124	0,205	0,248
	31-45	207	92	115	0,343	0,23
	46-99	31	16	15	0,059	0,03
BMI	0-23.09	80	7	73	0,026	0,146
	24-28.05	139	28	111	0,104	0,222
	28.06-32.05	137	51	86	0,190	0,172
	32.06-35.09	143	62	81	0,213	0,162
	36.01-39.05	94	39	55	0,145	0,11
	39.06-55	177	83	94	0,309	0,188

Sumber: (Maulidah et al, 2020)

Probabilitas prior yang digunakan yaitu probabilitas hasil seleksi atribut menggunakan partcile swarm optimization, probabilitas prior ini digunakan untuk menentukan kelas pada kasus baru yang terlebih dahulu dihitung probabilitas posterior nya.

4. PEMBAHASAN

4.1. Komparasi Model Algoritma Naive Bayes dan Particle Swarm Optimization

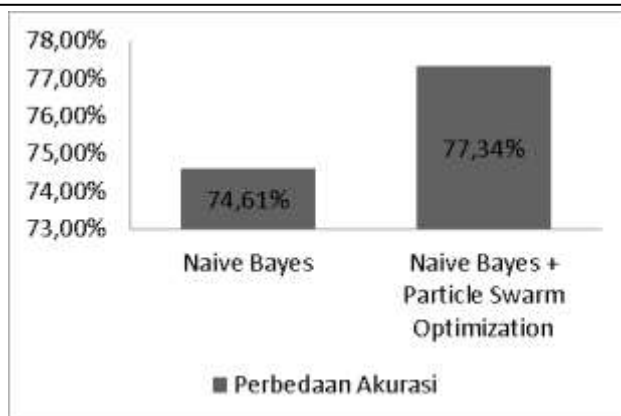
Hasil pengujian menyimpulkan bahwa Atribut yang mempengaruhi pengujian adalah 8 atribut. Hasil pengujian Naive Bayes tanpa seleksi atribut dibandingkan dengan model Naive Bayes dengan seleksi atribut menggunakan Particle Swarm Optimization dapat dilihat pada tabel berikut:

Tabel 5. Pengujian Algoritma klasifikasi Naive Bayes dan Naive Bayes+ Particle Swarm Optimization

	Accuracy
Naive Bayes	74,61%
Naive Bayes + Particle Swarm Optimization	77,34%

Sumber: (Maulidah et al, 2020)

Berikut perbedaan akurasi model Algoritma Naive Bayes dan model Algoritma Naive Bayes dengan Particle Swarm Optimization.



Sumber: (Maulidah et al, 2020)

Gambar 2 Perbedaan Akurasi Algoritma Naive Bayes dan Algoritma Naive Bayes dengan Particle Swarm Optimization.

Dari hasil pengujian diatas, dengan dilakukan evaluasi baik secara confusion matrix maupun ROC curve terbukti bahwa pengujian yang dilakukan Optimasi algoritma Naive Bayes menggunakan Particle Swarm Optimization memiliki nilai akurasi yang lebih tinggi dibanding hanya menggunakan algoritma Naive Bayes. Nilai akurasi untuk model algoritma Naive Bayes sebesar 74,61% dan nilai akurasi model Naive Bayes dengan seleksi atribut menggunakan Particle Swarm Optimization sebesar 77,34%. Berdasarkan nilai tersebut diperoleh selisih akurasi sebesar 2,73%.

5. KESIMPULAN

Dari hasil penelitian yang telah dilakukan untuk akurasi algoritma klasifikasi naive bayes sebesar 74,61%,sedangkan untuk akurasi algoritma klasifikasi dengan Particle Swarm Optimization sebesar 77,34% dengan selisih akurasi sebesar 2,73%. Sehingga dapat disimpulkan bahwa penerapan teknik optimasi Particle Swarm Optimization mampu menyeleksi atribut pada Algoritma Naive Bayes, dan dapat menghasilkan tingkat akurasi diagnosis penyakit diabetes yang lebih baik dibandingkan hanya menggunakan metode individual yaitu algoritma naive bayes.

DAFTAR PUSTAKA

- [1] KEMENKES RI, "Hari Diabetes Sedunia Tahun 2018," *Pus. Data dan Inf. Kementrian Kesehat. RI*, pp. 1–8, 2019.
- [2] A. Kantono, I. Y. Purbasari, and F. T. Anggraeny, "Penerapan pruning pada algoritma c5.0 untuk mendiagnosis penyakit diabetes melitus 1," no. September, pp. 184–189, 2019.
- [3] E. Y. Darmayanti, D. S. Budi, and A. B. Fitra, "Particle Swarm Optimization Untuk Optimasi Bobot Extreme Learning Machine Dalam Memprediksi Produksi Gula Kristal Putih Pabrik Gula," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 5096–5104, 2018.
- [4] Bustami, "Penerapan Algoritma Naive Bayes Untuk Nasabah Asuransi," *J. Inform.*, vol. 8, no. 1, pp. 884–898, 2014.