

Prediksi Penyakit Diabetes Dengan *Naïve Bayes*

Yuni Wardana¹, Devni Prima Sari^{1,2}

¹ Program Studi Matematika, Fakultas Matematika Ilmu Pengetahuan dan Alam, Universitas Negeri Padang (UNP)

² Data Analytics, Mathematical Modelling, and Forecasting Research Group, Universitas Negeri Padang, Indonesia

Article Info

Article history:

Received August 11, 2023

Revised August 18, 2023

Accepted September 20, 2023

Keywords:

Diabetes

Naïve Bayes

Prediction

Kata Kunci:

Diabetes

Naïve Bayes

Prediksi

ABSTRACT

Diabetes is one of the top ranked diseases for non-communicable diseases which is top cause of death worldwide. Every year 442 million people worldwide have diabetes and 1.6 million people die due to diabetes. With increasing cases of diabetes every year, this detection may need to be done before diabetes occurs. So a research was conducted on diabetes prediction using the *Naïve Bayes* method. According to the *Naïve Bayes* model, people with diabetes have a 28% chance of developing diabetes and a 72% chance of not having diabetes. And this research achieved 94% accuracy, 95% precision, and 98% recall.

ABSTRAK

Diabetes adalah salah satu penyakit yang menduduki peringkat teratas untuk golongan penyakit tidak menular penyebab teratas kematian di seluruh dunia. Setiap tahunnya 442 juta jiwa di seluruh dunia mengalami diabetes dan 1,6 juta jiwa mengalami kematian yang disebabkan oleh diabetes. Dengan peningkatan kasus diabetes setiap tahunnya, deteksi sedini mungkin perlu dilakukan sebelum diabetes terjadi. Maka dilakukan penelitian tentang prediksi diabetes menggunakan metode *Naïve Bayes*. Menurut model *Naïve Bayes*, penderita diabetes memiliki peluang 28% untuk terjangkit diabetes sedangkan 72% untuk tidak menderita diabetes. Penelitian ini mencapai tingkat akurasi 94%, presisi 95%, dan recall 98%.



Corresponding Author:

Yuni Wardana

Program Studi Matematika, Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,

Universitas Negeri Padang, Jl. Prof. Dr. Hamka, Air Tawar barat, Padang Utara, Padang, Indonesia. Kode Pos: 25131

Email: yuniwardana12@gmail.com

1. PENDAHULUAN

Masalah utama yang yang menyerang masyarakat Indonesia saat ini adalah PTM atau yang juga disebut dengan penyakit tidak menular. Hal ini ditandai dengan penyakit menular yang cenderung menurun menjadi PTM yang meningkat secara global dan dan menjadi 10 besar penyebab kematian terbanyak secara nasional, contohnya penyakit diabetes melitus dan penyakit metabolik [1]. Diabetes adalah penyakit jangka panjang dan berbahaya. Penyakit ini dapat berdampak serius pada organ dan jaringan tubuh. Dampak berat yang diterima penderita diabetes adalah masalah keuangan dan kehidupan sosialnya [2]. Diabetes adalah penyakit metabolik yang mengakibatkan sekresi insulin yang memadai, penurunan aktivitas insulin, atau keduanya yang ditandai dengan glukosa yang tinggi (*hiperglikemia*) [3].

Pengaturan gula normal merupakan kaitan yang erat dengan diabetes. Kalenjer pancreas yang memproduksi hormone insulin disebabkan oleh meningkatnya kadar gula darah. Komplikasi sering terjadi ketika seseorang menderita diabetes. *Stroke* dan penyakit jantung merupakan komplikasi yang mematikan dan sering terjadi. Hal ini terjadi karena glukosa yang terus meningkat, mengakibatkan terjadinya kerusakan saraf, pembuluh darah dan struktur internal lainnya. Gula yang terdapat pada dinding pembuluh menyebabkan penebalan pembuluh darah. Hal ini mengakibatkan kurangnya aliran darah, terutama yang menuju saraf dan kulit [4].

Perubahan gaya hidup masyarakat menjadi kurang sehat mengakibatkan meningkat jumlah penderita diabetes setiap tahunnya. Pada tahun 2015, *World Health Organization* (WHO) juga meliris fakta penting mengenai diabetes, yaitu di seluruh dunia terdapat 415 juta orang menderita diabetes dan diprediksi pada tahun 2040 diprediksi dapat mencapai 642 juta orang [5]. Pada tahun 2019, terdapat 4,2 juta kasus kematian karena diabetes dengan 463 juta kasus diabetes di seluruh dunia [6]. Menurut IDF, 578 juta kasus diabetes akan terjadi saat tahun 2030 dan pada saat tahun 2045 meningkat menjadi 700 juta kasus [7].

Dengan peningkatan kasus diabetes setiap tahunnya, deteksi sedini mungkin perlu dilakukan sebelum diabetes terjadi. Hal ini perhatian peneliti untuk melakukan beberapa analisis data terkait prediksi diabetes. Metode yang sering digunakan untuk analisis data adalah regresi, klasifikasi, dan clustering. Penelitian sebelumnya yang terkait dengan analisis data diabetes menggunakan metode regresi adalah menggunakan metode *Regresi Logistik* [8] [9], *Regresi Cox* [10], *Regresi Polinomial* dan *Regresi Eksponensial* [11]. Metode clustering juga digunakan dalam penelitian yang berkaitan dengan prediksi diabetes, diantaranya menggunakan *K-means Clustering* [12] [13], dan *Fuzzy dan C-means Clustering* [14] [15]. Kemudian penggunaan metode klasifikasi juga digunakan dalam analisis data diabetes antara lain PCA dan LDA [16] [17], *Bayesian Network*, dan *Support Vector Machines* [18] [19].

Pada penelitian ini, penulis menggunakan metode klasifikasi untuk memprediksi diabetes dengan metode *Naïve Bayes*. *Naïve Bayes* digunakan untuk mendapatkan prediksi probabilistik dari suatu data sehingga hasilnya berupa keputusan berdasarkan data training sebelumnya [20]. *Naïve Bayes* juga merupakan metode klasifikasi yang menggunakan asumsi independen menggunakan model probabilitas sederhana dengan teorema Bayes yang diajukan oleh ilmuwan Inggris Thomas Bayes. Asumsi bebas berarti bahwa keberadaan atribut tertentu dalam suatu kelas tidak mempengaruhi keberadaan atribut lainnya [21]. Keunggulan *Naïve Bayes* adalah efisien, komputasi cepat, dan tidak membutuhkan banyak data. Oleh karena itu, penelitian ini akan memprediksi diabetes menggunakan *Naïve Bayes*.

2. METODE

Naïve Bayes adalah algoritma yang memanfaatkan teorema bayes dan menggunakan asumsi bahwa nilai antar variabel saling bebas satu sama lain pada nilai keluaran untuk mendapatkan klasifikasi. Dengan demikian diasumsikan bahwa ada atau tidaknya suatu variabel tertentu tidak berhubungan dengan variabel lainnya. Rumus teorema bayes adalah



$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1)$$

Dengan

$P(A|B)$ = Peluang A pada kondisi B

$P(B|A)$ = Peluang B pada kondisi A

$P(A)$ = Peluang A

$P(B)$ = Peluang B

Teorema Bayes dapat ditulis secara ekivalen dengan persamaan 1.

$$Posterior = \frac{likelihood \cdot prior}{evidence} \quad (2)$$

Posterior merupakan peluang data yang diprediksi, *likelihood* merupakan peluang atribut data dengan kondisi A, *prior* adalah peluang kelas A, dan *evidence* merupakan karakteristik sampel secara keseluruhan. Menghitung peluang suatu data diawali dengan mencari *likelihood*. *Likelihood* yang didapat akan dikali dengan peluang dari masing-masing kelas. Untuk mengklasifikasikan data baru maka digunakan hasil dari proses tersebut.

Untuk membandingkan hasil klasifikasi yang dilakukan maka digunakan matriks konfusi. Matriks konfusi merupakan sebuah matriks yang membandingkan hasil klasifikasi oleh sistem dengan hasil klasifikasi sebenarnya [22]. *Confusion matrix* memperlihatkan beberapa informasi penting, yaitu akurasi, presisi, dan *recall* setiap kelas tujuan. Pada pengukuran kinerja menggunakan matriks konfusi.

Tabel 1 Matriks Konfusi

		Actual Value	
		1 (Positive)	0 (Negative)
Prediction Value	1 (Positive)	True Positif (TP)	False Positive (FP)
	0 (Negative)	False Negative (FN)	True Negative (TN)

Tabel 1 menunjukkan gambaran hasil klasifikasi dengan 4 istilah. Ini adalah *True Positive* (TP), *False Positive* (FP), *True Negative* (TN) dan *False Negative* (FN). Jumlah data negatif yang terdeteksi dengan benar adalah TN, sedangkan data negatif tetapi terdeteksi dengan positif adalah FP. FN adalah kebalikan dari TP, jadi datanya positif tetapi ternyata datanya negatif.

Deskripsi seberapa akurat sistem dapat mengklasifikasikan data sebagai fungsi dari nilai akurasi. Nilai akurasi juga dapat dijelaskan dengan membandingkan data yang diklasifikasikan dengan benar dengan seluruh kumpulan data. Deskripsi akurasi antara data yang diminta dan hasil prediksi yang diberikan oleh model sebagai fungsi dari nilai presisi. Sedangkan keberhasilan model dalam mengambil informasi adalah fungsi *recall*.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (3)$$

$$Presisi = \frac{TP}{FP + TP} * 100\% \quad (4)$$

$$Recall = \frac{TP}{FN + TP} * 100\% \quad (5)$$

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini diperoleh dari Institut Nasional Diabetes dan Penyakit Pencernaan dan Ginjal [23]. Dataset berisi 390 data, terdiri dari 60 data pasien diabetes positif (15,4%) dan 330 data pasien negatif diabetes (84,6%), dengan delapan variabel. Dengan

penggunaan data dibagi menjadi 2 yaitu 70% data training dan 30% data testing. Rincian variabel dijelaskan pada Tabel 2.

Tabel 2. Detail Variabel

Nama Variabel	Deskripsi dari Variabel	State	Deskripsi dari State
Diabetes (X_8)	Hasil diagnosa penyakit	1	Tidak Diabetes
		2	Diabetes
Kolesterol (X_1)	Atribut ini menentukan banyak kolesterol yang ada dalam tubuh	1	$X_1 < 200$
		2	$200 \leq X_1 < 300$
		3	$X_1 \geq 300$
Jenis Kelamin (X_2)	Atribut ini menentukan jenis kelamin	1	Wanita
		2	Pria
Usia (X_3)	Atribut ini menentukan usia	1	$X_3 < 40$
		2	$X_3 \geq 40$
BMI (X_4)	Atribut ini menentukan indeks massa tubuh	1	$X_4 < 20$
		2	$20 \leq X_4 < 35$
		3	$35 \leq X_4 < 50$
		4	$X_4 \geq 50$
Glukosa (X_5)	Atribut ini menentukan tingkat gula dalam tubuh	1	$X_5 < 100$
		2	$100 \leq X_5 < 150$
		3	$X_5 \geq 150$
<i>High Density Lipoprotein</i> (X_6)	Atribut ini menentukan banyak kolesterol baik yang ada di tubuh	1	$X_6 < 50$
<i>Waist Hip Ratio</i> (X_7)	Atribut ini membandingkan lingkaran pinggang dengan lingkaran pinggul	1	$X_7 < 0.75$
		2	$X_7 \geq 0.75$

Kolesterol (X_1) merupakan lemak yang diproduksi organ hati atau diperoleh dengan mengonsumsi produk hewani. Kadar kolesterol (X_1) yang tinggi dapat menyebabkan peningkatan kadar asam urat dan gula darah. Jenis kelamin (X_2) resiko terjadinya diabetes bisa didapatkan oleh pria ataupun wanita. Resiko lebih besar dialami oleh wanita untuk diserang penyakit diabetes dibandingkan laki-laki. Adapun pengaruh usia (X_3) adalah ketika semakin tua usia akan menyebabkan hormon insulin tidak dapat bekerja secara maksimal yang akan menyebabkan tingginya kadar gula darah. BMI (X_4) merupakan faktor resiko penyakit diabetes, ketika seseorang mengalami obesitas maka tubuhnya akan lebih sulit dalam menggunakan insulin.

Glukosa (X_5) merupakan penyebab utama penyakit diabetes, kadar gula darah yang tinggi akan meningkatkan peluang diabetes. Pada penderita diabetes ketika kadar gula tinggi akan memicu kadar kolesterol baik atau HDL (X_6) menurun. *Waist Hip Ratio* (X_7) merupakan perbandingan lingkaran pinggang dan panggul yang digunakan sebagai metode untuk mengetahui obesitas abdominal.

3.1. Perhitungan Probabilitas Diabetes

Penggunaan data pada tahap ini adalah data training 70%. Data training berjumlah 273 data diketahui bahwa banyak orang yang tidak menderita diabetes adalah 231 orang dan yang menderita diabetes adalah 42 orang. Probabilitas diabetes pada 273 data dapat diperhatikan pada Tabel 3

Tabel 3. Probabilitas Diabetes		
Variabel	State	Probabilitas
X_8	1	0,846
	2	0,154



3.2. Perhitungan Probabilitas Gejala Diabetes

Pada tahap ini dilakukan penghitungan probabilitas *likelihood*. Perhitungan *likelihood* dilakukan pada 273 dataset menggunakan gejala penyakit diabetes sebagai vektor. Pemilihan atribut prediksi diabetes adalah $X_1, X_2, X_3, X_4, X_5, X_6, X_7$. Tabel 4 hingga Tabel 10 menunjukkan perhitungan probabilitas atribut $X_1, X_2, X_3, X_4, X_5, X_6$, dan X_7 yang dipengaruhi oleh X_8 .

Tabel 4 Probabilitas ($X_1|X_8$)

Variabel	($X_1 X_8=1$)	($X_1 X_8=2$)	$P(X_1 X_8=1)$	$P(X_1 X_8=2)$
$X_1 < 200$	108	12	0,466	0,289
$200 \leq X_1 < 300$	113	27	0,487	0,622
$X_1 \geq 300$	10	3	0,047	0,089

Probabilitas ($X_1|X_8$) merupakan peluang kadar kolesterol pada penderita diabetes. Pada Tabel 4 diperlihatkan bahwa peluang kadar kolesterol $X_1 < 200$ pada seseorang dengan diabetes sebesar 0,289, peluang kadar kolesterol $200 \leq X_1 < 300$ pada seseorang dengan diabetes sebesar 0,622, dan peluang kadar kolesterol $X_1 \geq 300$ pada seseorang dengan diabetes sebesar 0,089. Jadi, peluang kadar kolesterol $200 \leq X_1 < 300$ pada seseorang dengan diabetes lebih besar daripada kadar kolesterol lainnya.

Tabel 5 Probabilitas ($X_2|X_8$)

Variabel	($X_2 X_8=1$)	($X_2 X_8=2$)	$P(X_2 X_8=1)$	$P(X_2 X_8=2)$
Wanita	146	22	0,631	0,523
Pria	85	20	0,369	0,477

Probabilitas ($X_2|X_8$) merupakan peluang jenis kelamin pada penderita diabetes. Pada Tabel 5 diperlihatkan bahwa peluang pasien wanita penderita diabetes adalah 0,523, dan peluang pasien pria penderita diabetes adalah 0,477. Jadi, peluang seseorang wanita menderita diabetes lebih besar daripada seorang pria.

Tabel 6 Probabilitas ($X_3|X_8$)

Variabel	($X_3 X_8=1$)	($X_3 X_8=2$)	$P(X_3 X_8=1)$	$P(X_3 X_8=2)$
$X_3 < 40$	94	3	0,408	0,091
$X_3 \geq 40$	137	39	0,592	0,909

Probabilitas ($X_3|X_8$) merupakan peluang usia penderita diabetes. Pada Tabel 6 diperlihatkan bahwa peluang usia dibawah 40 tahun pada pasien diabetes adalah 0,091 dan peluang usia diatas 40 tahun pada pasien diabetes adalah 0,909. Jadi, seseorang yang berusia diatas 40 tahun menderita diabetes memiliki kemungkinan lebih besar.

Tabel 7 Probabilitas ($X_4|X_8$)

Variabel	($X_4 X_8=1$)	($X_4 X_8=2$)	$P(X_4 X_8=1)$	$P(X_4 X_8=2)$
$X_4 < 20$	13	1	0,060	0,024
$20 \leq X_4 < 35$	183	33	0,783	0,785
$35 \leq X_4 < 50$	34	7	0,149	0,167
$X_4 \geq 50$	1	1	0,009	0,024

Probabilitas ($X_4|X_8$) merupakan peluang BMI penderita diabetes. Pada Tabel 7 diperlihatkan bahwa peluang BMI $X_4 < 20$ pada seseorang dengan diabetes sebesar 0,024, peluang BMI $20 \leq X_4 < 35$ pada seseorang dengan diabetes sebesar 0,785, peluang BMI $35 \leq X_4 < 50$ pada seseorang dengan diabetes sebesar 0,167, dan peluang memiliki BMI $X_4 \geq 50$ pada seseorang dengan diabetes sebesar 0,024. Jadi, peluang BMI $20 \leq X_4 < 35$ pada seseorang dengan diabetes memiliki kemungkinan lebih besar

Tabel 8 Probabilitas ($X_5|X_8$)

Variabel	($X_5 X_8=1$)	($X_5 X_8=2$)	$P(X_5 X_8=1)$	$P(X_5 X_8=2)$
$X_5 < 100$	177	2	0,761	0,067
$100 \leq X_5 < 150$	48	10	0,209	0,244
$X_5 \geq 150$	6	30	0,030	0,689

Probabilitas ($X_5|X_8$) merupakan peluang kadar glukosa pada penderita diabetes. Pada Tabel 8 diperlihatkan bahwa peluang kadar glukosa $X_5 < 100$ pada seseorang dengan diabetes sebesar 0,067, peluang kadar glukosa $100 \leq X_5 < 150$ pada seseorang dengan diabetes sebesar 0,244, dan peluang kadar glukosa $X_5 \geq 150$ pada seseorang dengan diabetes sebesar 0,689. Jadi, kadar glukosa $X_5 \geq 150$ pada seseorang dengan diabetes memiliki kemungkinan lebih besar.

Tabel 9 Probabilitas ($X_6|X_8$)

Variabel	($X_6 X_8=1$)	($X_6 X_8=2$)	$P(X_6 X_8=1)$	$P(X_6 X_8=2)$
$X_6 < 50$	126	32	0,545	0,750
$X_6 \geq 50$	105	10	0,455	0,250

Probabilitas ($X_6|X_8$) merupakan peluang *high density lipoprotein* pada penderita diabetes. Pada Tabel 9 diperlihatkan bahwa peluang *high density lipoprotein* $X_6 < 50$ pada seseorang dengan diabetes sebesar 0,750 dan peluang *high density lipoprotein* $X_6 \geq 50$ pada seseorang dengan diabetes sebesar 0,250. Jadi, *high density lipoprotein* $X_6 < 50$ pada seseorang dengan diabetes memiliki kemungkinan lebih tinggi.

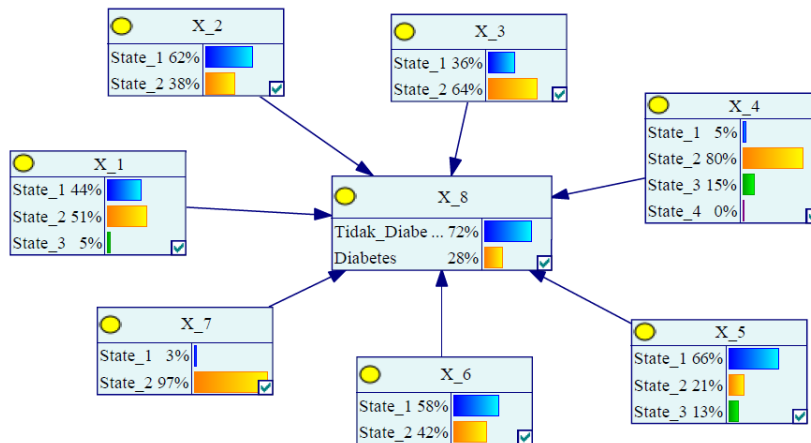
Tabel 10 Probabilitas ($X_7|X_8$)

Variabel	($X_7 X_8=1$)	($X_7 X_8=2$)	$P(X_7 X_8=1)$	$P(X_7 X_8=2)$
$X_7 < 0,75$	8	1	0,039	0,024
$X_7 \geq 0,75$	223	41	0,961	0,976

Probabilitas ($X_7|X_8$) merupakan peluang *waist hip ratio* pada penderita diabetes. Pada Tabel 10 diperlihatkan bahwa peluang *waist hip ratio* $X_7 < 0,75$ pada seseorang dengan diabetes sebesar 0,024 dan peluang *waist hip ratio* $X_7 \geq 0,75$ pada seseorang dengan diabetes sebesar 0,976. Jadi, *waist hip ratio* $X_7 \geq 0,75$ pada seseorang dengan diabetes memiliki kemungkinan lebih tinggi.

3.3. Perhitungan Probabilitas Prediksi Diabetes

Prediksi penderita diabetes dilakukan dengan mengalikan nilai probabilitas dari setiap atribut untuk mendapatkan nilai probabilitas prediksi penderita diabetes. Dengan asumsi nilai *evidence* selalu konsisten di semua kelas A, perhitungan *evidence* dapat dihilangkan. Perhitungan probabilitas *posterior* dilakukan dengan persamaan 1. Maka prediksi diabetes didapat $P(X_8 = 1) = 0,722$ dan $P(X_8 = 2) = 0,278$. Model *Naïve Bayes* untuk prediksi penyakit diabetes dapat dilihat pada Gambar 1



Gambar 1 Model *Naive Bayes* Untuk Diabetes

Pada model Gambar 1 dapat dilihat bahwa probabilitas kolesterol (X_1) adalah 44% untuk $X_1 < 200$, 51% untuk $200 \leq X_1 < 300$, dan 5% untuk $X_1 \geq 300$. Probabilitas jenis kelamin (X_2) wanita adalah 62% dan pria 38%. Probabilitas usia (X_3) adalah 36% untuk usia dibawah 40 dan 64% untuk yang diatas 40 tahun. Probabilitas BMI (X_4) adalah 5% untuk yang memiliki BMI $X_4 < 20$, 80% untuk BMI $20 \leq X_4 < 35$, 15% untuk BMI $35 \leq X_4 < 50$, dan 0% untuk BMI $X_4 \geq 50$. Probabilitas glukosa (X_5) adalah 66% untuk $X_5 < 100$, 21% untuk $100 \leq X_5 < 150$, dan 13% untuk $X_5 \geq 150$. Probabilitas *high density lipoprotein* (X_6) adalah 58% untuk $X_6 < 50$ dan 42% untuk $X_6 \geq 50$. Probabilitas *waist hip ratio* (X_7) adalah 3% untuk $X_7 < 0,75$ dan 97% untuk $X_7 \geq 0,75$. Dengan probabilitas $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ maka dengan metode *Naive Bayes* probabilitas terjadinya diabetes adalah 28% dan kemungkinan tidak terjadi diabetes adalah 72%

3.4. Validasi *Naive Bayes*

Evaluasi kinerja model *Naive Bayes* dilakukan dengan data testing sebanyak 117 data diabetes. Tabel keputusan digunakan pada tahap evaluasi model evaluasi kinerja *Naive Bayes*. Tabel keputusan untuk model prediksi penyakit diabetes dapat dilihat pada Tabel 11

Tabel 11 Tabel Konfusi

		<i>Actual Value</i>		Total
		Tidak Diabetes	Diabetes	
Prediction Value	Tidak Diabetes	97	5	102
	Diabetes	2	13	15
Total		99	18	117

Pada Tabel 11 dapat diketahui bahwa jumlah *true positive* adalah 97 yaitu prediksi terhadap pasien tidak diabetes yang benar mengalami *negative* diabetes berjumlah 97 orang. Jumlah *true negative* adalah 13 yaitu orang yang di prediksi diabetes memang benar diabetes. Jumlah *false positive* adalah 5 yaitu 5 orang yang di prediksi tidak diabetes ternyata diabetes. Jumlah *false negative* adalah 2 yaitu orang yang diprediksi diabetes ternyata tidak diabetes.

Hasil pengujian berupa matriks konfusi pasien dapat dianalisis menjadi akurasi, presisi dan *recall* dengan persamaan berturut-turut dari persamaan (3), (4.), (5) hasil pengujian disajikan pada Tabel 12 .

Tabel 12 Peforma Hasil Pengujian

Parameter	Nilai
Akurasi	94%
Presisi	95%
Recall	98%

Hasil dari model *Naïve Bayes* untuk prediksi diabetes memberikan hasil yang dapat diukur dalam nilai akurasi, presisi, dan *recall*, berturut-turut nilainya 94%, 95%, dan 98%

4. KESIMPULAN

Model *Naïve Bayes* yang diterapkan pada prediksi diabetes dapat menjelaskan hubungan antara diabetes dan gejalanya. Menurut model *Naïve Bayes*, peluang seseorang untuk terjangkit diabetes adalah 28% sedangkan peluang seseorang tidak terjangkit diabetes sebesar 72%. Penelitian ini mencapai tingkat akurasi 94%, presisi 95%, dan *recall* 98%.

REFERENSI

- [1] S. N. R. Toharin, M. Widya Hary Cahyati S.KM and d. I. Z. MH.Kes, "Hubungan Modifikasi Gaya Hidup Dan Kepatuhan Konsumsi Obat Antidiabetik Dengan Kadar Gula Darah Pada Penderita Diabetes Melitus Tipe 2 Di RS QIM Batang Tahun 2013," *UNNES JOURNAL OF PUBLIC HEALTH*, vol. 4, no. 2, 2015.
- [2] Y. Zhang, T. M. D. D Ph, S. E. Mann, Y. Chen, J. Martin, V. Moore, A. Baldwin, V. A. Reidel and W. W. Quick, "The Economic Costs of Undiagnosed Diabetes," *The Economic Cost Of Undiagnosed Diabetes 12*, 2009.
- [3] A. Bulu, T. D. Wahyuni and A. Sutriningsih, "Hubungan Antara Tingkat Kepatuhan Minum Obat Dengan Kadar Gula Darah Pada Pasien Diabetes Melitus Tipe II," *Nursing News: Jurnal Ilmiah Keperawatan*, 2019.
- [4] N. L. U. Sumadewi, *Isolasi Senyawa Tanin dan Uji Efek Hipoglikemik Ekstrak Kulit Batang Bungur (Lagerstroemia Speciosa Pers.) Terhadap Darah Mencit Yang diinduksi Aloksan*, Bali, Indonesia: Unud Library, 2011.
- [5] Eniarti, Sardjito Menyata, 2021.
- [6] N. Cho, J. Shaw, S. Karuranga, Y. Huang, J. d. R. Fernandes, A. Ohlogge and B. Malanda, "IDF Diabetes Atlas: Global Estimates Of Diabetes Prevalence For 2017 And Projections For 2015," *Diabetes Research and Clinical Practice*, pp. 271-281, 2018.
- [7] IDF, "International Diabetes Federation : Diabetes Atlas Ninth Edition," USA, 2019.
- [8] R. D. Joshi and C. K. Dhakal, "Predicting Type 2 Diabetes Using Logistic Regression and Machine Learning Approaches," *International Journal of Environmental Research and Public Health*, 2021.
- [9] C. Zhu, C. U. Idemudia and W. Feng, "Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques," *Informatics in Medicine Unlocked*, 2019.
- [10] R. Dankner, N. Agay, L. Olmer, H. Murad, L. K. Boker, R. D. Balicer and L. S. Freedman, "Metformin Treatment and Cancer Risk: Cox Regression Analysis with Time-Dependent Covariates of 320,000 Individuals with Incident Diabetes Mellitus," *Am J Epidemiol*, 2019.
- [11] A. Mukasheva, "The Prevalence of Diabetes in the Republic of Kazakhstan Based on Regression Analysis Methods," *International Journal of Health and Medical Sciences*, 2019.
- [12] N. Arora, A. Singh, M. Z. N. Al-Dabagh and S. K. Maitra, "A Novel Architecture for Diabetes Patients' Prediction Using K-Means Clustering and SVM," *Mathematical Problems in Engineering*, 2022.
- [13] M. Raihan, M. T. Islam, F. Farzana, M. G. M. Raju and H. S. Mondal, "An Empirical Study to Predict Diabetes Mellitus using K-Means and Hierarchical Clustering Techniques," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019.
- [14] A. Anggrawan and M. Mayadi, "Application of KNN Machine Learning and Fuzzy C-Means to Diagnose Diabetes," *Jurnal : Manajemen, Teknik Informatika dan Rekayasa Komputer*, 2023.
- [15] G. V. R, D. J. N and D. A. A. K. N, "Classification Of Gestational Diabetes Using Modified Fuzzy C Means Clustering And Machine Learning Technique," *Turkish Journal of Computer and Mathematics Education*, 2021.
- [16] M. Maniruzzaman, M. M. Islam, M. J. Rahman, M. A. M. Hasan and J. Shin, "Risk prediction of diabetic nephropathy using machine learning techniques: A pilot study with secondary data," *Diabetes and Metabolic Syndrome: Clinical Research and Review*, 2021.
- [17] S. K. Das, T. S. Bhattacharya, M. Ghosh and J. Chowdhury, "Probing blood plasma samples for the detection of diabetes using SERS aided by PCA and LDA multivariate data analyses," *New Journal of Chemistry*, 2021.
- [18] R. D. H. Devi, A. Bai and N. Nagarajan, "A novel hybrid approach for diagnosing diabetes mellitus using farthest first and support vector machine algorithms," *Obesity Medicine*, 2020.
- [19] N. Mohan and V. Jain, "Performance Analysis of Support Vector Machine in Diabetes Prediction Diabetes Prediction," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2020.
- [20] Wisdayani, D. Selvy and dkk, "Perbandingan Algoritma K-Nearest Neighbor dan *Naïve Bayes* untuk Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas di Kabupaten Pati Jawa Tengah," *Skripsi Universitas Muhammadiyah Semarang*, 2019.
- [21] M. Windariti, "Perbandingan Kinerja Algoritma *Naïve Bayes* dan Bayesian Network dalam Klasifikasi Masa Studi Mahasiswa," *Prosiding Seminar Nasional Aplikasi Sains & Teknologi*, 2018.



-
- [22] E. Prasetyo, *Data Mining : Konsep dan Aplikasi Menggunakan Matlab*, Yogyakarta: Andi Offset, 2021.
- [23] "Predict diabetes based on diagnostic measures | Kaggle," <https://www.kaggle.com/datasets/houcembenmansour/predict-diabetes-based-on-diagnostic-measures> (accessed 02 April 2023).
- [24] Y. Mukhtar, U. M. Yunusa and G. A.M, "A Modern Overview on Diabetes Mellitus: A Choric Endocrine Disorder," *European Journal of Biologi*, pp. 1-14, 2020.
- [25] Y. Yu, O. A. Arah, Z. Liew, S. Cnattingius, J. Olsen, H. T. Sørensen, G. Qin and J. Li, "Maternal diabetes during pregnancy and early onset of cardiovascular disease in offspring: population based cohort study with 40 years of follow-up," *BMJ*, 2019.