

KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA NAIVE BAYES

Devi Nurul Anisa¹, Jumanto²

^{1,2}Universitas Negeri Semarang

Program Studi Teknik Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Negeri Semarang

e-mail: devianisa1219@students.unnes.ac.id, jumanto@mail.unnes.ac.id

ABSTRAK

Penyakit diabetes adalah suatu penyakit gangguan metabolik yang di tandai oleh tingginya gula darah yang melebihi nilai normal. Terdapat banyak faktor yang menjadi penyebab penyakit diabetes, faktor-faktor tersebut diantaranya seperti faktor keturunan, berat badan, usia, dan faktor lainnya. Banyak yang tidak menyadari bahwa dirinya terkena penyakit diabetes, sehingga angka kematian yang disebabkan oleh penyakit diabetes ini semakin banyak dan setiap tahunnya diperkirakan akan terus meningkat angka kasus kematiannya. Maka dari itu penelitian ini mencoba menerapkan suatu metode klasifikasi untuk memprediksi apakah seseorang terkena diabetes atau tidak. Dataset yang digunakan pada penelitian ini merupakan data yang di dapatkan dari data Kaggle, yaitu Predict diabetes based on diagnostic measure. Metode klasifikasi yang digunakan yaitu dengan menerapkan algoritma Naive Bayes yang mampu menghasilkan akurasi yang baik. Hasil dari penelitian ini di dapati nilai akurasi 92%. Hasil ini lebih baik dibanding dengan penelitian sebelumnya yang menggunakan k-nearest neighbor (KNN) dengan tingkat akurasi sebesar 91%.

Kata Kunci: *klasifikasi, diabetes, naive bayes*

1. PENDAHULUAN

Diabetes mellitus atau lebih sering disebut diabetes merupakan penyakit metabolis yang kronis yang mana penyakit diabetes terjadi karena kurangnya jumlah insulin yang cukup disebabkan ketidakmampuan pankreas dalam tubuh untuk memproduksi hormon insulin. Kekurangan insulin dapat mengakibatkan gula darah di dalam tubuh mengalami jumlah yang berlebihan, kondisi ini dapat menyebabkan komplikasi pada ginjal dan organ tubuh lainnya [1]. Insulin merupakan hormon yang disekresi oleh pankreas dalam tubuh manusia [2]. Pasien didiagnosa menderita penyakit diabetes pada saat kadar gula darah dalam darahnya melebihi nilai normal [3].

Penyakit diabetes ini merupakan penyakit yang cukup berbahaya. Penyakit ini menyebabkan komplikasi jangka panjang yang berkembang secara bertahap. Semakin lama penderita penyakit diabetes akan semakin tidak terkontrol gula darahnya, dan semakin tinggi risiko komplikasinya. Akhirnya, komplikasi diabetes dapat melumpuhkan dan bahkan mengancam jiwa [4]. Komplikasi yang sering terjadi dan mematikan akibat diabetes adalah serangan jantung dan stroke. Faktor kematian yang terjadi karena kenaikan gula darah secara terus menerus dan mengakibatkan rusaknya pembuluh darah, saraf dan struktur internal lainnya [5]. Penyakit diabetes melitus adalah penyakit yang memiliki komplikasi tinggi, indentifikasi awal merupakan satu-satunya cara untuk menghindari komplikasi [6].

Menurut Organisasi Kesehatan Dunia (WHO), diabetes menyebabkan 4,6 juta kematian pada tahun 2011 [7]. Pada tahun 2012, 2,2 juta kematian disebabkan oleh glukosa darah tinggi dan 1,6 juta kematian terjadi pada tahun 2016 disebabkan oleh diabetes [8]. Menurut International Diabetes Federation yang mengidap penyakit diabetes pada tahun 2015 sebanyak 415 juta jiwa, dan diperkirakan meningkat sebanyak 227 juta jiwa atau menjadi 642 juta jiwa pada tahun 2040 pasien yang mengidap penyakit diabetes [9]. Selain itu, pengeluaran biaya kesehatan untuk Diabetes Mellitus telah mencapai 465 miliar USD [10].

Guna menyikapi masalah ini, perlu melakukan deteksi dini penyakit diabetes. Hal ini dilakukan agar dapat melakukan pencegahan terhadap penyakit diabetes. Dapat dikembangkan suatu sistem untuk memprediksi penyakit dengan memanfaatkan berbagai metode. Salah satu metode yang dapat digunakan yaitu metode data mining dengan prinsip klasifikasi [11]. Data mining adalah sebuah metode untuk melakukan proses penambangan data yang menghasilkan sebuah output (keluaran) berupa pengetahuan. Terdapat beberapa bagian yang berkaitan dengan data mining, antara lain analisis klaster, analisis asosiasi, deteksi anomali dan model prediksi. Untuk model prediksi terdiri dari 2 jenis, yaitu klasifikasi dan regresi [12].

Klasifikasi merupakan sebuah proses algoritma untuk pencarian pola dengan membangun model klasifikasi berdasarkan variabel kelas prediktor dan variabel kelas target [13]. Beberapa algoritma yang digunakan untuk proses klasifikasi, diantaranya adalah K-Nearest Neighbor (knn) [14], Naive Bayes (nb) [15], Support Vector Machine (svm) [16], Neural Network (nn) [17], serta Ada Boost Classifier (abc) [18].

Algoritma Naive Bayes adalah algoritma pembelajaran untuk masalah klasifikasi yang terutama digunakan untuk klasifikasi teks yang melibatkan kumpulan data pelatihan berdimensi tinggi [19]. Klasifikasi menggunakan

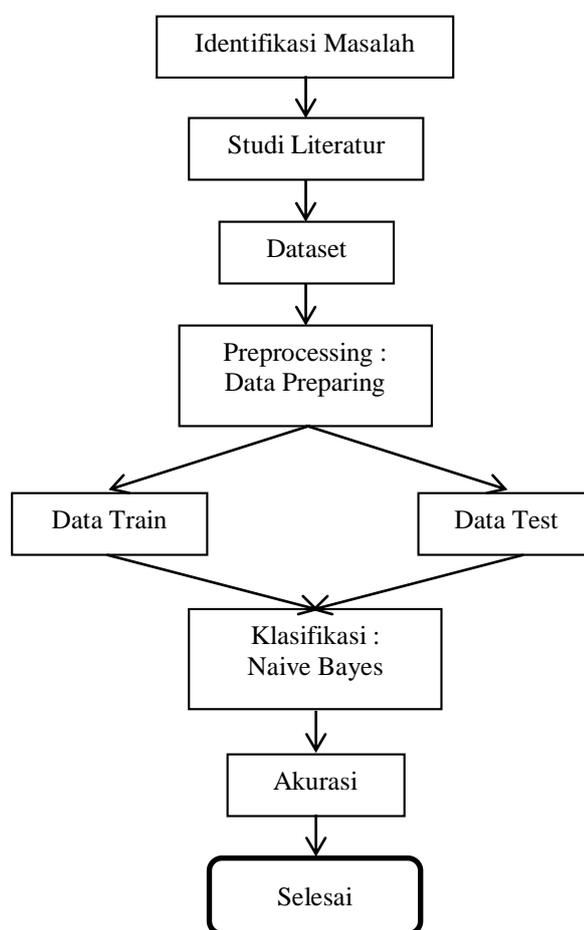
Naive Bayes dapat menghasilkan akurasi tinggi dan cepat ketika diaplikasikan pada data yang benar serta mudah dalam penerapannya [5].

Terdapat beberapa penelitian mengenai klasifikasi diabetes yang telah dilakukan sebelumnya, Purnama dkk. [12] menggunakan klasifikasi Neural Network pada dataset yang didapat dari Machine Learning Repository UCI dan didapatkan akurasi sebesar 80%. Sedangkan penelitian yang dilakukan Ridwan [14] terhadap data yang juga didapat dari Machine Learning Repository UCI yaitu Early stage diabetes risk prediction dataset dengan menggunakan Algoritma Naive Bayes menghasilkan akurasi sebesar 90,20%. Pada penelitian yang dilakukan Hassan dkk. [8] menggunakan 3 klasifikasi pada 1 data, yaitu dengan menggunakan klasifikasi Support Vector Machine (SVM), K-Nearest Neighbour (KNN), dan Decision Tree terhadap data Pima Indian Diabetic Dataset dan memperoleh akurasi 90,23 untuk klasifikasi menggunakan Support Vector Machine (SVM), akurasi 75,97% untuk klasifikasi menggunakan K-Nearest Neighbour (KNN), dan akurasi 75,32% untuk klasifikasi menggunakan Decision Tree.

Pada penelitian ini diimplementasikan algoritma Naive Bayes untuk memprediksi penyakit diabetes dengan menggunakan dataset yang sama dengan penelitian sebelumnya yaitu dataset yang diperoleh dari situs web Kaggle, diharapkan dengan menggunakan algoritma Naive Bayes dapat meningkatkan nilai akurasi yang dihasilkan.

2. METODE PENELITIAN

Penelitian yang dilakukan terdiri dari beberapa langkah, mulai dari identifikasi masalah, studi literatur, pengambilan dataset, preprocessing data hingga klasifikasi, sebagai mana ditunjukkan pada Gambar 1.



Gambar 1. Flowchart Penelitian

2.1 Identifikasi Masalah

Diabetes merupakan penyakit yang cukup berbahaya. Penyakit ini bisa menyebabkan komplikasi yang tinggi hingga dapat mengakibatkan kematian. Penyakit diabetes perlu dikenali sejak dini, sehingga dapat ditangani sesegera mungkin. Maka, perlu adanya klasifikasi untuk mengenali penyakit diabetes secara akurat. Algoritma Naive Bayes merupakan algoritma klasifikasi yang digunakan dalam penelitian ini. Naive Bayes memiliki

kemampuan yang cepat dalam membuat model, mempunyai kemampuan memprediksi dan juga menyediakan metode baru dalam mengeksplor dan memahami data.

2.2 Dataset

Pada penelitian ini, data didapatkan dari situs web Kaggle, Predict diabetes based on diagnostic measure. Dataset ini memiliki 15 variabel dan 390 data dengan 60 data didiagnosis penyakit diabetes dan 330 data didiagnosis sehat dari penyakit diabetes. Deskripsi variabel dataset diabetes dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi Variabel Diabetes

Variabel	Deskripsi
cholesterol	Kolesterol
glucose	Glukosa
hdl_chol	High-density lipoprotein yang merupakan kolesterol baik
chol_hdl_ratio	Perbandingan antar kolesterol di dalam darah
age	Umur
gender	Jenis kelamin
height	Tinggi badan
weight	Berat badan
bmi	Index masa tubuh atau proporsional tubuh
systolic_bp	Tekanan darah sistolik
diastolic_bp	Tekanan darah diastolik
waist	Lingkar perut
hip	Lingkar panggul
waist_hip_ratio	Perbandingan lingkar perut dan lingkar panggul
diabetes	Suatu penyakit metabolik yang diakibatkan oleh meningkatnya kadar glukosa atau gula darah.

2.3 Preprocessing

Sebelum dilakukan proses klasifikasi dengan menggunakan Naive Bayes, data terlebih dahulu dilakukan preprocessing. Preprocessing data dilakukan dengan menghilangkan beberapa variabel yang sekiranya tidak terlalu dibutuhkan. Selanjutnya, melakukan skaling terhadap data untuk normalisasi data. Setelah praproses, total 330 no diabetes dan 60 diabetes. Deskripsi detail dari dataset dan variabelnya di tunjukan pada Tabel 1. Dua variabel kelas digunakan untuk mengetahui apakah pasien memiliki risiko diabetes atau no diabetes. Kemudian dilakukan Data Splitting, yaitu membagi data menjadi data train dan data test, dengan presentase data train sebesar 60% dan data test 40%.

2.4 Klasifikasi

Pengklasifikasi Naive Bayes adalah pengklasifikasi sederhana berdaraskan teorema Bayesian dengan asumsi penentuan nasib sendiri yang kuat. Naive Bayes. Pengklasifikasi Naive Bayes menganggap bahwa ada (atau tidak adanya) fitur (atribut) tertentu dari suatu kelas tidak terkait dengan ada (atau tidak adanya) fitur lain ketika variabel kelas diberikan. Persamaan (1) ditunjukkan formula Naive Bayes.

$$p(C|F_1, \dots, F_n) = \frac{p(C)p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)} \quad (1)$$

Keuntungan menggunakan Naive Bayes adalah : (1) Cepat dan model sangat terukur. (2) Menyeimbangkan linear dengan jumlah prediktor dan baris. (3) Prosedur Naive Bayes adalah parallel. (4) Naive Bayes dapat digunakan untuk klasifikasi biner dan multiclass.

2.5 Akurasi

Akurasi merupakan alat ukur yang biasa digunakan untuk mengevaluasi kinerja model yang telah dibangun. Untuk menghitung akurasi ditunjukkan pada persamaan (2).

$$\text{Accuracy} = \frac{TP+TN}{P+N} \quad (2)$$

Dengan keterangan sebagai berikut : (1) TP atau True Positives adalah jumlah tuple positif yang dilabeli dengan benar oleh classifier. Tuple positif adalah tuple aktual yang berlabel positif, seperti tuple dengan label = 'Diabetes'. (2) TN atau True Negatives adalah jumlah tuple negatif yang dilabeli dengan benar oleh classifier. Tuple negatif adalah tuple aktual yang berlabel negatif, seperti tuple dengan label = 'No Diabetes'. (3) FP atau False Positives adalah jumlah tuple negatif yang salah dilabeli oleh classifier. Misalnya, sebuah tuple pasien yang berlabel = 'No Diabetes' akan tetapi oleh classifier dilabeli = 'Diabetes' (4) FN atau False Negatives adalah jumlah tuple positif yang salah dilabeli oleh classifier. Misalnya, sebuah tuple pasien yang berlabel = 'Diabetes' akan tetapi oleh classifier dilabeli = 'No Diabetes'. Empat istilah tersebut dapat digambarkan sebagai confusion matrix seperti yang diilustrasikan pada Tabel 2.

Tabel 2. Confusion Matrix

No	Kelas Aktual	Kelas Hasil Prediksi		
		Diabetes	No Diabetes	Jumlah
1	Diabetes	TP	FN	P
2	No Diabetes	FP	TN	N
3	Jumlah	P'	N'	P+N

4. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan 8 variabel atau atribut dan 1 label klasifikasi sebagaimana disajikan pada Tabel 3.

Tabel 3. Ada 8 variabel dataset gejala dan 1 label penentu klasifikasi.

No	Variabel	Value
1	Cholesterol	Normal under 200
2	Glucose	Normal 80-120
3	HDL Cholestrol	Normal more than 60
4	Age	1.20–35, 2.36–45, 3.46–55,4.56–65, 6.above 65
5	Gender	1. Female 2. Male
6	Weight	Sesuai indeks massa tubuh
7	Systolic BP	Normal 120
8	Distolic BP	Normal 80
9	Diabetes	1. Diabetes, 2. No Diabetes

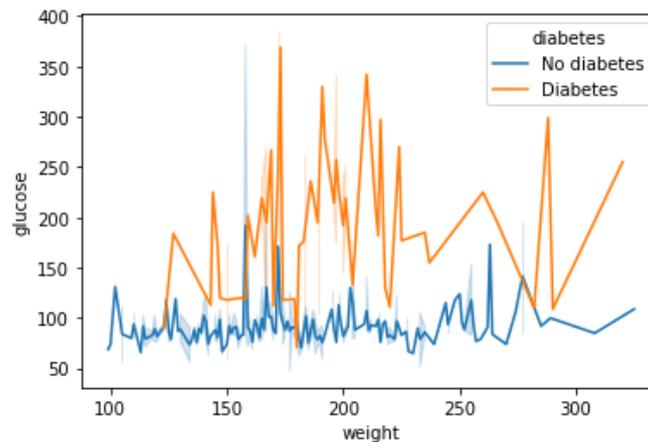
4.1 Preprocessing

Dalam proses preprocessing dilakukan dengan menghilangkan beberapa variabel yang tidak terlalu diperlukan, dari 16 variabel yaitu kolesterol, glucose, HDLkolesterol, kolesterol HDL ratio, age, gender, height, weight, bmi, systolic BP, diastolic BP, waist, HIP, waist HIP ratio, diabetes. Di dapat 9 variabel yang akan digunakan pada penelitian yaitu kolesterol, glucose, hdl kolesterol, age, gender, weight, systolic bp, distolic bp, diabetes. Data yang telah dipilih di tunjukan pada Tabel 3. Selanjutnya, melakukan skaling terhadap data untuk normalisasi data. Setelah praproses, total 330 no diabetes dan 60 diabetes. Kemudian dilakukan Data Splitting, yaitu membagi data menjadi data train dan data test, dengan presentase data train sebesar 60% dan data test 40% dan akan di klasifikasi oleh algoritma Naive Bayes.

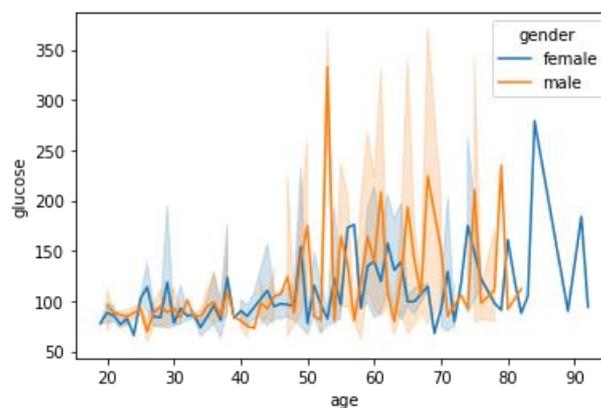
4.2 Klasifikasi Naive Bayes

Untuk menentukan pasien mempunyai penyakit diabetes atau tidak, pada penelitian ini telah dibuat rangkuman dari dataset yang digunakan. Gambar 2 ditunjukkan grafik pengaruh berat badan dan kadar gula terhadap pasien diabetes. Secara umum, Pasien diabetes memiliki kadar glukosa yang lebih tinggi dan berat badan yang lebih tinggi dibandingkan dengan yang non-diabetes. Gambar 3 ditunjukkan grafik pengaruh usia, jenis kelamin terhadap kadar gula seseorang yang berpotensi diabetes. Usia tidak secara langsung berhubungan tetapi

kadar glukosa yang lebih tinggi pada orang tua dapat menjadi penyebab Diabetes pada mereka, juga laki-laki usia 40 sampai 80 tahun memiliki kadar glukosa darah lebih tinggi daripada perempuan.

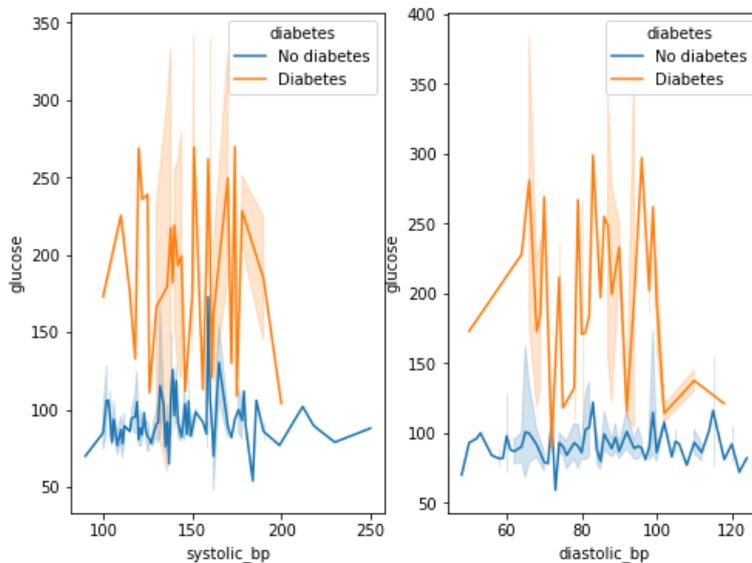


Gambar 2. Grafik pengaruh berat badan dan kadar gula

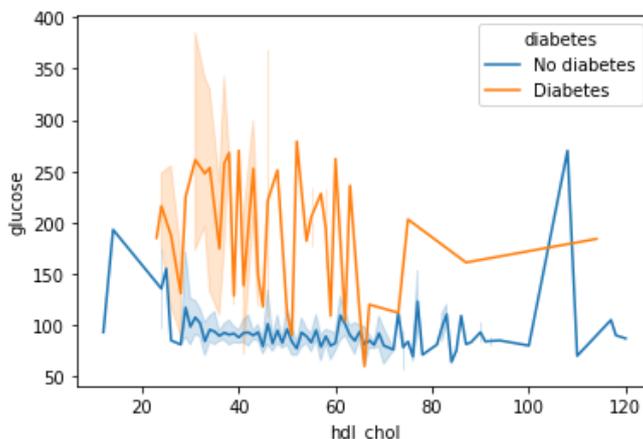


Gambar 3. Grafik pengaruh usia, jenis kelamin, dan kadar gula

Gambar 4 ditunjukkan grafik pengaruh tekanan darah terhadap pasien diabetes. Tekanan darah tidak berhubungan langsung dengan diabetes, karena pasien dengan tekanan darah tertinggi ditemukan bukan penderita diabetes. Gambar 5 ditunjukkan grafik pengaruh kolesterol HDL terhadap pasien diabetes. Pasien diabetes memiliki kolesterol HDL yang lebih rendah.

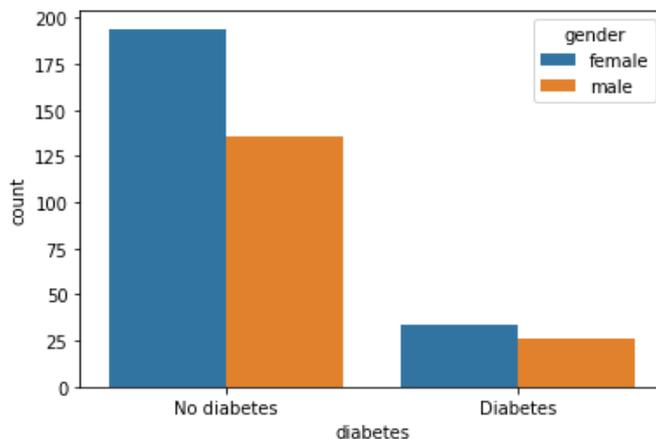


Gambar 4. Grafik pengaruh tekanan darah terhadap diabetes

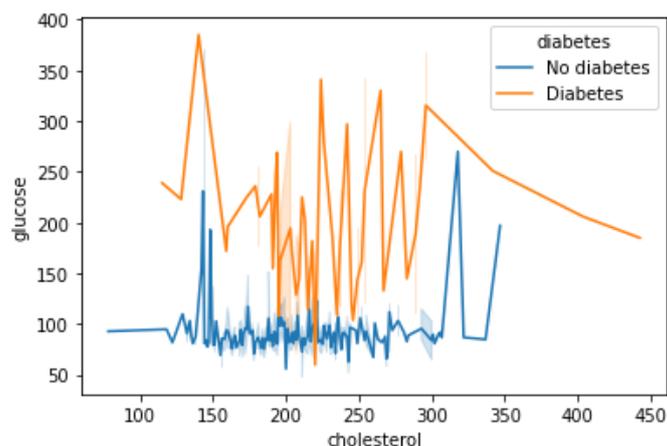


Gambar 5. Grafik pengaruh kolesterol HDL terhadap pasien diabetes

Gambar 6 ditunjukkan grafik perbandingan pasien perempuan dan laki-laki yang terkena penyakit diabetes. Perempuan yang menderita diabetes lebih banyak daripada laki-laki yang menderita diabetes. Gambar 7 ditunjukkan grafik pengaruh kolesterol terhadap penyakit diabetes. Kolesterol yang lebih tinggi terlihat pada pasien yang menderita diabetes.



Gambar 6. Grafik perbandingan pasien diabetes perempuan dan laki-laki



Gambar 7. Grafik pengaruh kolesterol terhadap penyakit diabetes

Seperti yang telah kita lihat pada rangkuman di atas juga, Glukosa berhubungan langsung dengan penyebab diabetes. Kolesterol baik atau HDL-kolesterol berhubungan negatif, berat badan, usia dan kolesterol juga terkait tetapi pada tingkat yang sangat kecil, pasien laki-laki memiliki kadar Glukosa yang lebih tinggi tetapi dari total pasien, pasien perempuan lebih banyak menderita diabetes daripada laki-laki, tetapi perbedaannya sangat kecil.

Adapun contoh data yang menunjukkan pasien diabetes dan no diabetes dapat dilihat pada Tabel 4. Pada dataset ada total 390 data. Diantaranya 330 adalah No Diabetes dan 60 adalah Diabetes.

Tabel 4. Data pasien diabetes dan no diabetes

No	Variabel	Value	
		1	2
1	Cholesterol	232	230
2	Glucose	184	112
3	HDL Cholesterol	114	64
4	Age	91	20
5	Gender	Female	Male
6	Weight	127	159
7	Systolic BP	170	100
8	Distolic BP	82	90
9	Diabetes	Diabetes	No Diabetes

4.3 Evaluasi

Model Klasifikasi yang telah dibuat selanjutnya dievaluasi menggunakan nilai akurasi yang dapat dihitung dengan menggunakan rumus yang direpresentasikan menggunakan confusion matrix sehingga diperoleh nilai akurasi pada proporsi data 60:40 adalah

Untuk data train menghasilkan confusion matrix sebagaimana ditunjukkan pada Tabel 5.

Tabel 5. Tabel Hasil Confusion Matrix Data Train

		Actual Values	
		Diabetes	No Diabetes
Predicted Values	Diabetes	24	10
	No Diabetes	8	192

$$\text{Confusion matrix} = \begin{bmatrix} 24 & 10 \\ 8 & 192 \end{bmatrix}$$

Sehingga nilai akurasi untuk data train adalah dengan TP 24, TN 192, P 32, dan N 202 dihasilkan nilai akurasi sebesar 92,3%.

Untuk data test menghasilkan confusion matrix sebagaimana ditunjukkan pada Tabel 6.

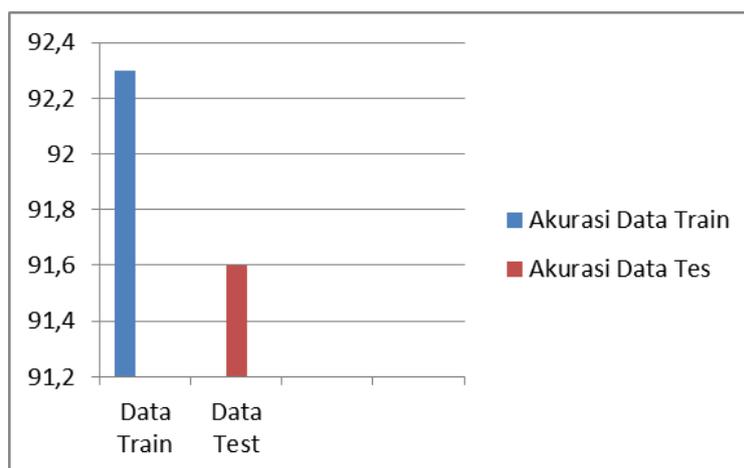
Tabel 6. Tabel Hasil Confusion Matrix Data Test

		Actual Values	
		Diabetes	No Diabetes
Predicted Values	Diabetes	19	7
	No Diabetes	6	124

$$\text{Confusion matrix} = \begin{bmatrix} 19 & 7 \\ 6 & 124 \end{bmatrix}$$

Sehingga nilai akurasi untuk data test adalah dengan TP 19, TN 124, P 25, dan N 231 dihasilkan nilai akurasi sebesar 91,6%.

Gambar 8 ditampilkan hasil dari evaluasi pengklasifikasian dengan Naive Bayes. Dari Gambar 8 didapatkan akurasi untuk mengklasifikasikan Dataset Diabetes yaitu 92,3% untuk Data Train dan 91,6% untuk Data Test.



Gambar 8. Akurasi Data Train dan Test dengan Naive Bayes

5. KESIMPULAN

Penelitian ini menggunakan Algoritma Naive Bayes untuk pengklasifikasian Dataset Predict diabetes based on diagnostic measure. Terdapat 390 data yang akan di klasifikasi. Dalam pengklasifikasian menggunakan 9 variabel yaitu kolesterol, glucose, hdl kolesterol, age, gender, weight, systolic bp, distolic bp, diabetes. Hasil Evaluasi klasifikasi ini menghasilkan akurasi untuk mengklasifikasikan Dataset Diabetes yaitu 92,3% untuk Data Train dan 91,6% untuk Data Test, artinya Klasifikasi permodelan Algoritma Naive Bayes terhadap Dataset Stage Diabetes sudah bagus akurasinya, tetapi perlu peningkatan akurasi dengan cara lain. Penelitian ini dapat dikembangkan lebih lanjut untuk mendapatkan akurasi yang lebih baik. Berikut saran yang dapat diberikan, yaitu : (1) Dapat menambahkan metode lain yang dapat di gabungkan dengan naive bayes. (2) Pengembangan dapat menggunakan algoritma data mining lain yang biasa digunakan, seperti K-means dan Support Vector Machine (SVM). (3) Diujikan menggunakan dataset lain.

DAFTAR PUSTAKA

- [1] V. K. Putri and F. I. Kurniadi, "Klasifikasi Diabetes Menggunakan Model Pembelajaran Ensemble Blending," 2018.
- [2] D. Z. Maliha, E. Santoso, and M. T. Furqon, "Penerapan Metode Neighbor Weighted K-Nearest Neighbor Dalam Klasifikasi Diabetes Mellitus," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2910–2915, 2019, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4817>.
- [3] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [4] A. W. Mucholladin, F. A. Bachtiar, and M. T. Furqon, "Klasifikasi Penyakit Diabetes menggunakan Metode Support Vector Machine," vol. 5, no. 2, 2021.
- [5] F. Fitriyani, "Prediksi Diabetes Menggunakan Algoritma Naive Bayes dan Greedy Forward Selection," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 61–69, 2021, doi: 10.25077/teknosi.v7i2.2021.61-69.
- [6] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1578–1585, 2018, doi: 10.1016/j.procs.2018.05.122.
- [7] K. M. Aamir, L. Sarfraz, M. Ramzan, M. Bilal, J. Shafi, and M. Attique, "A fuzzy rule-based system for classification of diabetes," *Sensors*, vol. 21, no. 23, 2021, doi: 10.3390/s21238095.
- [8] A. S. Hassan, I. Malaserene, and A. A. Leema, "Diabetes Mellitus Prediction using Classification Techniques," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 5, pp. 2080–2084, 2020, doi: 10.35940/ijitee.e2692.039520.
- [9] R. P. Fadhillah, R. Rahma, A. Sepharni, R. Mufidah, B. N. Sari, and A. Pangestu, "Klasifikasi Penyakit Diabetes Mellitus Berdasarkan Faktor-Faktor Penyebab Diabetes Menggunakan Algoritma C4.5," *J. Ilm. Penelit. dan Pembelajaran Inform.*, vol. 07, no. 04, pp. 1265–1270, 2022, [Online]. Available: <https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/424/338>.
- [10] J. J. Purnama, S. Rahayu, S. Nurdiani, T. Haryanti, and N. A. Mayangky, "Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Diabetes," *Indones. J. Comput. Inf. Technol.*, vol. 5, no. 1, pp. 1–7, 2019, doi: 10.33480/pilar.v15i2.601.
- [11] M. S. Efendi and H. A. Wibawa, "Prediksi Penyakit Diabetes Menggunakan Algoritma ID3 dengan Pemilihan Atribut Terbaik (Diabetes Prediction using ID3 Algorithm with Best Attribute Selection)," vol. VI, pp. 29–35, 2018.
- [12] L. U. Khasanah, Y. N. Nasution, F. Deny, and T. Amijaya, "Klasifikasi Penyakit Diabetes Melitus Menggunakan Algoritma Naive Bayes Classifier," *J. Ilm. Mat.*, vol. 1, no. 1, pp. 41–50, 2022.
- [13] G. Abdurrahman, "Jurnal Sistem dan Teknologi Informasi Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier," vol. 7, no. 1, 2022.
- [14] M. Cubillos, S. Wøhlk, and J. N. Wulff, "A bi-objective k-nearest-neighbors-based imputation method for multilevel data," *Expert Syst. Appl.*, vol. 204, p. 117298, Oct. 2022, doi: 10.1016/j.eswa.2022.117298.
- [15] Y. F. Safri, R. Arifudin, and M. A. Muslim, "K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor," *Sci. J. Informatics*, vol. 5, no. 1, p. 18, May 2018, doi: 10.15294/sji.v5i1.12057.
- [16] S. Sulistiana and M. A. Muslim, "Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy," *J. Soft Comput. Explor.*, vol. 1, no. 1, Sep. 2020, doi: 10.52465/jossex.v1i1.3.
- [17] J. Jumanto, M. F. Mardiansyah, R. N. Pratama, M. F. Al Hakim, and B. Rawat, "Optimization of breast cancer classification using feature selection on neural network," *J. Soft Comput. Explor.*, vol. 3, no. 2, Sep. 2022, doi: 10.52465/jossex.v3i2.78.
- [18] X. Huang, Z. Li, Y. Jin, and W. Zhang, "Fair-AdaBoost: Extending AdaBoost method to achieve fair

- classification,” *Expert Syst. Appl.*, vol. 202, p. 117240, Sep. 2022, doi: 10.1016/j.eswa.2022.117240.
- [19] A. Ridwan, “Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.