Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Des 17, 2024 | Revised Des 24, 2024 |
Accepted Jan 02, 2025

Hal. 11-20
Vol. 7; No. 1
Februari 2025

ANALISIS PERBANDINGAN ALGORITMA KLASIFIKASI UNTUK IDENTIFIKASI DIABETES DENGAN MENGGUNAKAN METODE RANDOM FOREST DAN NAÏVE BAYES

Muhammad Rafli Zuhri[1*], Kusrini[2], Dhani Ariatmanto[3]
[1,2,3]Magister Teknik Informatika, Universitas Amikom Yogyakarta
email: mrzrafli29@students.amikom.ac.id[1*]

Abstrak: Penanganan penyakit diabetes menjadi penting karena komplikasi yang dapat terjadi jika tak ditanggulangi dengan benar. Oleh karena itu, pengembangan metode yang efektif dalam mendiagnosis penyakit diabetes pada perempuan menjadi sangat penting. Klasifikasi merupakan salah satu metode yang dapat digunakan untuk mengidentifikasi diabetes. Random Forest dan Naïve Bayes merupakan dua algoritma klasifikasi yang populer. Random Forest adalah metode kompleks yang didasarkan pada penggabungan beberapa pohon keputusan untuk mengdapatkan prediksi yang lebih akurat(Supriyadi et al., 2020). Sedangkan Naïve Bayes merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan denegan asumsi antar variabel penjelas saling bebas (independen). Tujuan dari penelitian ini untuk menganalisis perbandingan antara algoritma Naïve Bayes dan Random Foret dalam mengidentifikasi penyakit diabetes. Hasil penelitian digunakan data sebanyak 70% sebagai data training dan 30% sebagai data testing dari keseluruhan 768 data keseluruhan didapatkan bahwa metode random forest dapat memprediksi penyakit diabetes dengan tingkat persentase sebesar 94% dan tingkat persentase naïve bayes sebesar 78% Berdasarkan hasil penelitian didapatkan metode random forest memiliki tingkat persentase akurasi lebih tinggi dibandingkan metode naïve bayes dengan tingkat persentase 94% sedangkan naïve bayes dengan tingkat persentase 78% sehingga dapat disimpulkan bahwa metode random forest merupakan metode terbaik dalam mengindentifikasi penyakit diabetes dibandingkan metode naïve bayes

Kata Kunci : diabetes, diagnosa, klasifikasi, naïve bayes, random forest

PENDAHULUAN

Diabetes adalah suatu penyakit metabolik yang diakibatkan oleh meningkatnya kadar glukosa atau gula darah. Gula darah sangat vital bagi kesehatan karena merupakan sumber energi yang penting bagi sel-sel dan jaringan. Jika tidak dikelola dengan baik, diabetes dapat menyebabkan terjadinya berbagai komplikasi, seperti penyakit jantung koroner, stroke, obesitas, serta gangguan pada mata, ginjal, dan saraf [1].

Diabetes mellitus (DM), menurut definisi *World Health Organization* (WHO), Karena gejalanya yang mirip dengan kondisi sakit biasa, banyak orang yang tidak menyadari bahwa mereka mengidap penyakit diabetes dan bahkan sudah mengarah pada komplikasi[2], Penyakit diabetes kini menyerang manusia tanpa mengenal usia. Bahkan lebih dari 1,2 juta anak-anak dan remaja di dunia terkena penyakit diabetes. Penyakit diabetes pun masih masuk ke daftar penyakit paling mematikan di dunia[3] Berdasarkan data *International Diabetes Federation* (IDF), Indonesia berada dalam status waspada diabetes dan Indonesia sendiri berada di urutan ke-7 dari 10 negara dengan jumlah penderita diabetes terbanyak di dunia [4].

Penanganan penyakit diabetes menjadi penting karena komplikasi yang dapat terjadi jika tak ditanggulangi dengan benar. Oleh karena itu, pengembangan metode yang efektif dalam mendiagnosis penyakit diabetes pada perempuan menjadi sangat penting [5]. Banyak faktor yang mempengaruhi orang menderita diabetes, beberapa diantaranya yaitu tekanan darah tinggi, kadar gula berlebih, berat badan, riwayat keturunan diabetes, usia, jumlah kehamilan seseorang, ketebalan lipatan kulit, dan jumlah kadar insulin dalam tubuh[6]. Berbagai metode telah digunakan untuk mengidentifikasi diabetes, termasuk pemeriksaan fisik, tes darah, dan tes urine, namun hasil pemeriksaan membutuhkan waktu yang lama untuk mengetahui hasilnya sehingga dibutuhkan teknologi untuk memberikan kemudahan setiap orang dalam menerima informasi secara cepat.

Salah satu contohnya yaitu penerapan teknologi dalam bidang kesehatan karena membutuhkan peralatan yang mampu memberikan diagnosa suatu penyakit dengan beberapa pertimbangan sehingga teknik data mining bisa digunakan untuk memberikan prediksi ataupun mengklasifikasikan suatu penyakit berdasarkan himpunan data yang terdapat pada rumah sakit maupun layanan kesehatan lain [7].

Klasifikasi merupakan salah satu metode yang dapat digunakan untuk mengidentifikasi diabetes. Algoritma klasifikasi ini dapat menganalisis data pasien, seperti usia, jenis kelamin, riwayat kesehatan, dan hasil tes, untuk memprediksi apakah pasien tersebut memiliki diabetes atau tidak [8]. Algoritma klasifikasi terdiri dari 5 yaitu *Neural Network, K-Nearest Neighbors, Decision Tree, Random Forest. Random Forest* dan *Naïve Bayes* merupakan dua algoritma klasifikasi yang populer. *Random Forest* adalah metode kompleks yang didasarkan pada penggabungan beberapa pohon keputusan untuk mengdapatkan prediksi yang lebih akurat [9] Sedangkan *Naïve Bayes* merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan denegan asumsi antar variabel penjelas saling bebas (independen). Pada algoritma ini pembelajaran lebih ditekankan pada pengestimasian probabilitas. Keuntungan algoritma *Naïve Bayes* adalah tingkat nilai error yang didapat lebih rendah ketika dataset berjumlah besar, selain itu akurasi *Naïve Bayes* dan kecepatannya lebih tinggi pada saat diaplikasikan kedalam dataset yang jumlahnya lebih besar. Tujuan dari penelitian ini untuk menganalisis perbandingan antara algoritma *Naïve Bayes*

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Des 17, 2024 | Revised Des 24, 2024 |
Accepted Jan 02, 2025
Hal. 11-20
Vol. 7; No. 1
Februari 2025

dan *Random Foret* dalam mengidentifikasi penyakit diabetes. Alasan penelitian ini menggunakan dua metode tersebut karena dua metode ini merupakan metode yang paling umum digunakan dalam melakukan penelitian dengan tujuan penelitian untuk membandingkan kedua metode tersebut serta mencari metode mana yang memiliki akurasi yang lebih baik. Penelitian ini diharapkan dapat memberikan informasi yang bermanfaat tentang kinerja *Random Forest* dan *Naïve Bayes* dalam mengidentifikasi diabetes. Hasil penelitian ini dapat membantu para dokter dan peneliti dalam mengembangkan metode yang lebih efektif untuk mengidentifikasi diabetes.

TINJAUAN PUSTAKA

Penelitian tentang klasifikasi diagonosis penyakit Stroke menggunakan algoritma *Random Forest.* Penelitian ini bertujuan untuk dengan menggunakan Metode *Random Forest* menjadi pilihan tepat dalam melakukan preprosessing data dalam mengidentifikasi gejala awal. Hasil model penyesuaian menghasilkan 96% skor pelatihan dan dari tabel hasil *precision, recall, F1-score,* dan *accuracy* Yang mendapatkan hasil akurasi sebesar 0.95 atau 95%, serta hasil akhir dari AUC sebesar 0.80 yang menunjukan hasil model tersebut termasuk ke dalam klasifikasi baik [10].

Penelitian tentang klasifikasi penyakit diabetes menggunakan *Naïve Bayes*. Tujuan penelitian ini adalah agar mempermudah dunia medis khususnya dokter ahli menentukan suatu klasifikasi Diabetes Mellitus kepada pasien. Hasil Penelitian *Performancevector*: *Accuracy*: 35.00% *Confusionmatrix*: *True*: Dirawat Pulang Dirawat: 3 9 Pulang: 4 4 *Precision*: 50.00% (Positive Class: Pulang) *Confusionmatrix*: *True*: Dirawat Pulang Dirawat : 3 9 Pulang: 4 4 *Recall:* 30.77% (*Positive Class*: Pulang) *Confusionmatrix: True*: Dirawat Pulang Dirawat : 3 9 Pulang : 4 4 [11]. Penelitian tentang analisis perbandingan akurasi untuk klasifikasi diabetes dengan menggunakan algoritma *Naïve Bayes* dan C4.5. Tujuan penelitian ini dilakukan untuk mengetahui hasil perbandingan nilai performa algoritma *Naïve Bayes* dan C4.5 dengan 7 skenario berbeda pada klasifikasi penyakit diabetes yang akan diuji performa *accuracy, precision,* dan *recall*[12]. Hasil penelitian kami menunjukkan bahwa algoritma C4.5 (skenario 4) memiliki hasil yang baik dalam klasifikasi penyakit diabetes dibandingkan algoritma *Naïve Bayes* (skenario 2) dimana performa algoritma C4.5 memiliki *accuracy* 99.03%, *precision* 100%, dan *recall* 98.18%.

Penelitian tentang perbandingan algoritma klasifikasi Metabolik Sindrom dengan menggunakan algoritma ¬*Naïve Bayes* dan *K-Nearest Neighbors*. Tujuan penelitian ini untuk menentukan algoritma model mana yang memiliki nilai akurasi, *pressi*, dan *recall* yang lebih tinggi. Penelitian ini juga melakukan mengevaluasi tingkat akurasi dari tiga spliting data. Hasil penelitian ini menunjukkan bahwa penggunaan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 79%, sedangkan akurasi tertinggi dari algoritma *K-Nearest Neighbors* (KNN) adalah 82%. Kesimpulannya, dari hasil penelitian ini menunjukkan bahwa algoritma K-NN dengan pembagian data 50:50 lebih efektif dalam memprediksi dan mengklasifikasikan sindrom metabolic [13].

Penelitian tentang perbandingan implementasi Machine Learning untuk klasifikasi penyakit diabetes dengan menggunakan metode *KNN, Naïve Bayes*, dan *Logistik Regression*[14]. Tujuan penelian ini memberikan pemahaman mendalam tentang potensi dan kecocokan metode klasifikasi tertentu dalam menangani permasalahan klasifikasi data diabetes melalui pemilihan metode ekstraksi fitur yang tepat, pembagian data yang representatif, dan evaluasi kinerja yang cermat.Tujuan penelitian ini terfokus pada menganalisis kinerja tiga metode klasifikasi utama, yaitu *K-Nearest Neighbor* (KNN), *Naive Bayes*, dan *Logistic Regression*, dalam konteks pengklasifikasian data diabetes[15]. Tujuan dari penelitian ini untuk menganalisis perbandingan antara algoritma *Naïve Bayes* dan *Random Forest* dalam mengidentifikasi penyakit diabetes. Alasan penelitian ini menggunakan dua metode tersebut karena dua metode ini merupakan metode yang paling umum digunakan dalam melakukan penelitian dengan tujuan penelitian untuk membandingkan kedua metode tersebut serta mencari metode mana yang memiliki akurasi yang lebih baik. Hasil penelitian ini dapat membantu para dokter dan peneliti dalam mengembangkan metode yang lebih efektif untuk mengidentifikasi diabetes.

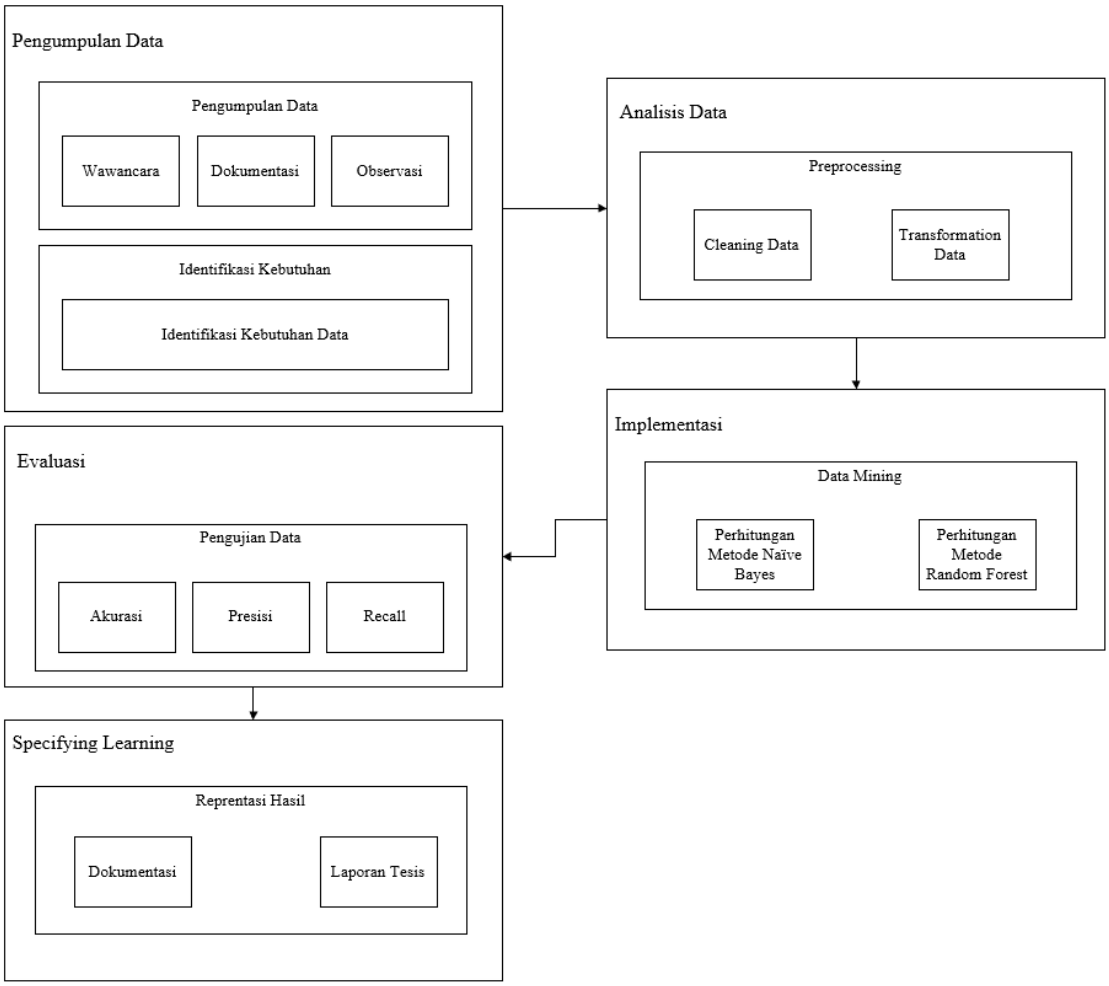
METODE

Penelitian yang dilakukan bersifat deskriptif karena menggambarkan atau menganalisis secara rinci karakteristik suatu fenomena tanpa mengambil keputusan sebab-akibat. Penelitian deskriptif dimana data dijelaskan dalam bentuk angka dan tabel atau diagram. Setelah data diolah akan dilakukan analisis dengan pendekatan kuantitatif yang dijelaskan dengan hasil perhitungan angka dan tabel atau diagram. Pada penelitian ini menggunakan pendekatan kuantitatif yang nantinya hasil dari penelitian ini merupakan informasi-informasi berupa angka dan diagram hasil dari eksperimen penggabungan dua metode yang dilakukan.

Metode analisis data yang dilakukan dalam penelitian ini menggunakan teknik data mining dengan menerapkan algoritma *Naïve Bayes* dan *Random Forest* untuk dilakukan komparasi akurasi dalam klasifikasi penyakit diabetes..

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Des 17, 2024 | Revised Des 24, 2024 |
Accepted Jan 02, 2025

Hal. 11-20
Vol. 7; No. 1
Februari 2025



Gambar 1. Alur Penelitian

1. Diagnosing

Pada tahap pengumpulan data ini melakukan pengumpulan data dengan mengambil data sampel pasien berupa adalah data usia, indeks massa tubuh (BMI), tekanan darah, dan hasil tes laboratorium terkait gula darah yang kemudian dikelompokkan ke dalam pasien yang terdiagnosa diabetes dan tidak terdiagnosa untuk kemudian dilakukan identifikasi kebutuhan data yang sesuai dengan sistem yang akan dibangun.

2. Analisis Data

Cleaning Data, Untuk tahapan cleaning data dilakukan proses pembersihan data dengan menghilangkan missing value atau bisa disebut data tidak berisi atau kosong (null), serta data yang tidak lengkap. Kemudian data yang akan dibersihkan akan melalui proses *cleaning* dengan melalui cara *Replace Missing Values* untuk mengisi nilai rata-rata atribut tertentu disetiap daerah yang kosong yang mengacu pada atributnya.

Transformation Data, Tahap ini dilakukan proses transformasi atau normalisasi data kedalam format yang dapat dikelola oleh sistem. Dengan cara normalisasi menggunakan tools rapid miner dan mengubah format data awal sesuai dengan kategori dikarenakan analisis asosiasi hanya bisa menerima input data kategorikal, transformasi pada kolom data kontinu kemudian dilakukan pemisahan data dengan membagi data training dan data testing dengan persentase 70%:30% dari total data keseluruhan hasil dari pre-processing.

Implementasi, Melakukan perhitungan random forest dan perhitungan *Naive Bayes* untuk mencari akurasi perbandingan dari kedua metode tersebut.

Evaluating, Melakukan pengujian hasil dari implementasi *Naïve Bayes* dan *Random forest* dengan menggunakan metode *Confusion Matrix* untuk mencari nilai akurasi, *presisi*, *recall* dan F1-*score* untuk melihat tingkat persentase dari kedua metode.

Specifying Learning, Pada tahap ini dilakukan proses dokumentasi dan publikasi thesis berisi hasil peneltian yang sudah diterapkan.

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Des 17, 2024 | Revised Des 24, 2024 |
Accepted Jan 02, 2025

Hal. 11-20
Vol. 7; No. 1
Februari 2025

HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Pada Table 1. Pada tahap ini dilakukan pengambilan sampel data pasien seperti pada gambar 4 dibawah ini yang di ambil dari kaagle yang berisi 8 variabel dan 1 class hasil yang terdiri dari 768 data. Detail atribut pada penelitian ini seperti pada table 1 dibawah ini.

Table 1. Fitur Data Set

No	Fitur	
	Atribut	Deskripsi
2	Kehamilan	Menunjukkan nilai dari kehamilan
3	Glukosa	Menunjukkan level glukosa pada darah
4	Tekanan darah	Menunjukkan nilai dari hasil pengukuran tekanan darah
5	Ketebalan Kulit	Menunjukkan nilai dari ketebalan kulit
6	Insulin	Menunjukkan nilai insulin pada darah
7	BMI	Menunjukkan hasil pengukuran berat badan
8	Riwayat diabetes	Menunjukkan nilai dari riwayat diabetes
9	Umur	Menunjukkan umur dari pasien

Table 2. Kelas Dataset

No	**Fitur**	
	Atribut	Deskripsi
1	Hasil	Menunjukkan variabel klasifikasi ketika 0 artinya tidak menderita diabetes dan 1 berarti menderita diabetes

Pada Table 2. Data penelitian secara keseluruhan dapat dilihat pada Table 3. dibawah ini

Table 3. Data Penelitian

No	Kehamilan	Glukosa	Tekanan Darah	Ketebalan Kulit	Insulin	BMI	Riwayat Diabetes	Umur	Hasil
1	6	148	72	35	0	33,6	0,627	50	1
2	1	85	66	29	0	26,6	0,351	31	0
3	8	183	64	0	0	23,3	0,672	32	1
4	0	89	66	23	94	28,1	0,167	21	0
5	0	137	40	35	168	43,1	2,268	33	1
..	……	….	…	….	….	…	…….	…	…
764	10	101	76	48	180	32,9	0,171	63	0
765	2	122	70	27	0	36,8	0,340	27	0
766	5	121	72	23	112	26,2	0,245	30	0
767	1	126	60	0	0	30,1	0,349	47	1
768	1	93	70	31	0	30,4	0,315	23	0

Pada gambar Table 3 merupakan data yang peneliti gunakan dalam penelitian yang terdiri dari 768 data, pada data tersebut terdiri dari data kehamilan, glukosa, tekanan darah, ketebalan kulit, insulin, BMI, riwayat diabetes dan umur sedangkan hasil yaitu variabel klasifikasi dimana 0 jika tidak mengidap diabetes dan 1 jika mengidap diabetes.

2. Analisis Data

a. Cleaning Data

Pada tahap ini dilakukan pembersihan data dengan menghilangkan missing value atau bisa disebut data tidak berisi atau kosong (null), serta data yang duplikat.

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Des 17, 2024 | Revised Des 24, 2024 |
Accepted Jan 02, 2025

Hal. 11-20
Vol. 7; No. 1
Februari 2025

```
Jumlah missing value per kolom:
 Kehamilan           0
Glukosa              0
Tekanan Darah        0
Ketebalan Kulit      0
Insulin              0
BMI                  0
Riwayat Diabetes     0
Umur                 0
Hasil                0
```

Gambar 2. Hasil Missing Values

Pada gambar 3, merupakan hasil missing values dan duplikat data, pada gambar diatas terlihat bahwa hasilnya menunjukkan dataset yang akan dianalisis tidak memiliki data yang duplikat sehingga tidak ada data yang dibuang/dihapus.

b. Transformasi Data

Pada tahap ini dilakukan mengubah data mentah menjadi data desimal untuk mempermudah dan meningkatkan kualitas data untuk analisis yang lebih akurat. Hasil data transformasi data dapat dilihat pada tabel 4.4 dibawah ini.

Table 4. Data sebelum Normalisasi

No	Kehamilan	Glukosa	Tekanan Darah	Ketebalan Kulit	Insulin	BMI	Riwayat Diabetes	Umur	Hasil
1	6	148	72	35	0	33,6	0,627	50	1
2	1	85	66	29	0	26,6	0,351	31	0
3	8	183	64	0	0	23,3	0,672	32	1
4	0	89	66	23	94	28,1	0,167	21	0
5	0	137	40	35	168	43,1	2,268	33	1
..	……	….	…	….	….	…	…….	…	…
764	10	101	76	48	180	32,9	0,171	63	0
765	2	122	70	27	0	36,8	0,340	27	0
766	5	121	72	23	112	26,2	0,245	30	0
767	1	126	60	0	0	30,1	0,349	47	1
768	1	93	70	31	0	30,4	0,315	23	0

Pada Table 4. adalah data sebelum dilakukan normalisasi dan Table 5. merupakan data setelah dilakukan normalisasi data, tahap ini dilakukan dengan mengubah angka numerik ke angka desimal untuk melakukan pemerataan data agar kualitas data merata untuk menghasilkan analisis data yang akurat.

Table 5. Data sesudah Normalisasi

No	Kehamilan	Glukosa	Tekanan Darah	Ketebalan Kulit	Insulin	BMI	Riwayat Diabetes	Umur	Hasil
1	0,352	0,743	0,590	0.353	0	0,500	0,234	0.483	1
2	0,058	0,427	0,540	0,292	0	0,396	0,116	0,116	0
3	0,470	0,919	0,524	0	0	0,347	0,253	0,183	1
4	0,058	0,447	0,540	0,232	0,111	0,418	0,038	0	0
5	0	0,688	0,327	0,353	0,198	0,642	0,943	0,200	1
..	……	….	…	….	….	…	…….	…	…
764	0,588	0,507	0,622	0,484	0,212	0,490	0,039	0,700	0
765	0,117	0,613	0,537	0,272	0	0,548	0,111	0,100	0
766	0,294	0,608	0,590	0,232	0,132	0,390	0,071	0,150	0
767	0,058	0,633	0,491	0	0	0,448	0,115	0,433	1
768	0,058	0,467	0,573	0,313	0	0,453	0,101	0,033	0

c. Split Data

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Des 17, 2024 | Revised Des 24, 2024 |
Accepted Jan 02, 2025

Hal. 11-20
Vol. 7; No. 1
Februari 2025

Pada penelitian ini akan dilakukan proses split data menjadi 2 bagian yaitu data training dan data testing. Variasi pembagian data akan dilakukan menjadi beberapa bagian antara lain dengan rasio 70% data training dan 30% data testing, 80% data training dan 20% data testing, 90% data training dan 10% data testing. Variasi pembagian data ini merupakan upaya eksperimen untuk mengetahui teknik split data yang memiliki perfoma terbaik. Jumlah komposisi pembagian data seperti pada tabel 6. dibawah ini.

Table 6. Jumlah Komposisi

No	Training	Testing	Jumlah Data Training	Jumlah Data Testing
1	70%	30%	537	231
2	80%	20%	614	154
3	90%	10%	691	77

3. Data Mining

Hasil prediksi ini adalah hasil yang didapat dari split data menggunakan Naïve Bayes Classifier dan Random Forest. Pada Table 7 menampilkan hasil split data menggunakan Naïve Bayes Classifier dan pada Table 8 menampilkan split data menggunakan Random Forest. Dari hasil prediksi kita dapat memperoleh nilai akurasi, F1, dan presisi seperti yang terlihat di Table 7 dan Table 8. Nilai akurasi diperoleh dari persentase jumlah prediksi benar (Diabetes dan Tidak diabetes) dibanding dengan jumlah data test secara keseluruhan. Untuk F1 didapatkan dari nilai rata rata antara nilai dari presisi dan nilai recall dimana nilai recall merupakan perbandingan antara jumlah diprediksikan bernilai positif dengan banyak data positif yang memang benar positif. Terakhir, untuk nilai presisi diperoleh dari jumlah prediksi benar positif dibandingkan dengan jumlah hasil yang diprediksi positif.

a. Naïve Bayes

Pada tahap ini akan dibuatkan hasil prediksi data training dan data testing menggunakan *naïve bayes* dengan perbandingan 70% data training dan 30% data testing dari keseluruhan data.

Table 6. Hasil Naive Bayes Classifier

No	Split Data	*Naïve Bayes Classifier*		
		Akurasi	F1	Presisi
1	70% : 30%	76%	58%	67%
2	80% : 20%	79%	64%	67%
3	90% : 10%	81%	72%	70%

b. Random Forest

Table 7. Hasil Random Forest

No	Split Data	Random Forest		
		Akurasi	F1	Presisi
1	70% : 30%	76%	58%	66%
2	80% : 20%	74%	55%	59%
3	90% : 10%	79%	68%	71%

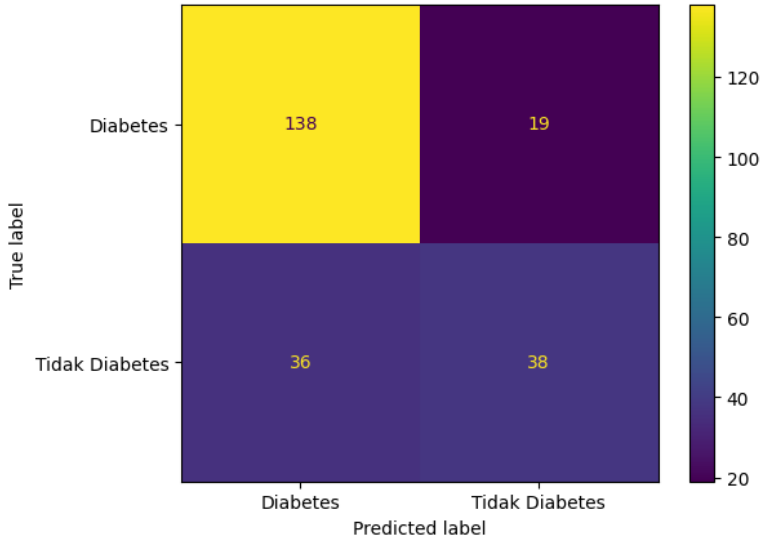
4. Evaluasi

Pada tahap ini dilakukan penilaian akurasi dari kedua metode untuk mengetahui metode terbaik dalam mengidentifikasi diabetes. Adapun hasil evaluasi dari pemodelan *Naïve Bayes Classifier* dan *Random Forest.*

a. Naïve Bayes

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Des 17, 2024 | Revised Des 24, 2024 |
Accepted Jan 02, 2025

Hal. 11-20
Vol. 7; No. 1
Februari 2025



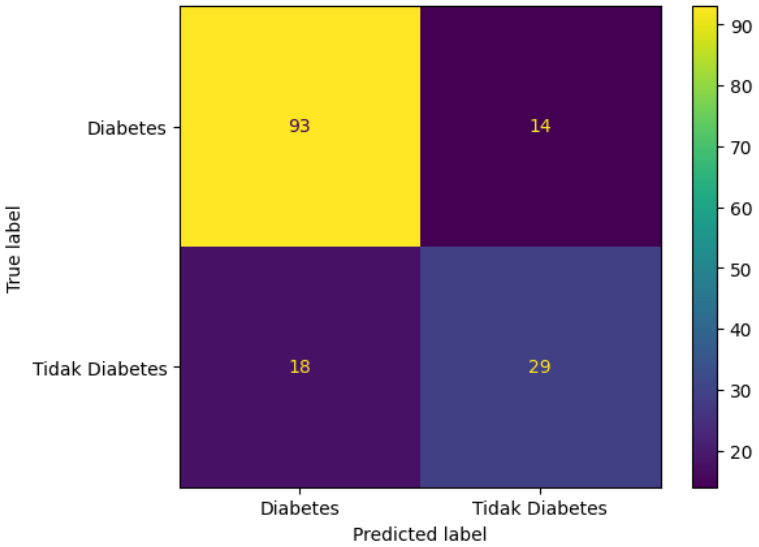
Gambar 3. ***Confusion Matrix Naïve Bayes 70%:30%***

Pada gambar 3. merupakan hasil dari confusion matrix dari naïve bayes dengan perbandingan data training dan data testing sebesar 70%:30%. Sehingga didapatkan hasil akurasi, presisi, recall dan F1-score sebagai berikut:

Classification Report

	precision	recall	f1-score	support
0.0	0.79	0.88	0.83	157
1.0	0.67	0.51	0.58	74
accuracy			0.76	231
macro avg	0.73	0.70	0.71	231
weighted avg	0.75	0.76	0.75	231

Gambar 4. Hasil Confusion Matrix 70%:30%.



Gambar 5. Confusion Matrix 80%:20%

Pada gambar 5. merupakan hasil dari *confusion matrix* dari *naïve bayes* dengan perbandingan data training dan data testing sebesar 80%:20%. Sehingga didapatkan hasil akurasi, presisi, recall dan F1-score sebagai berikut:

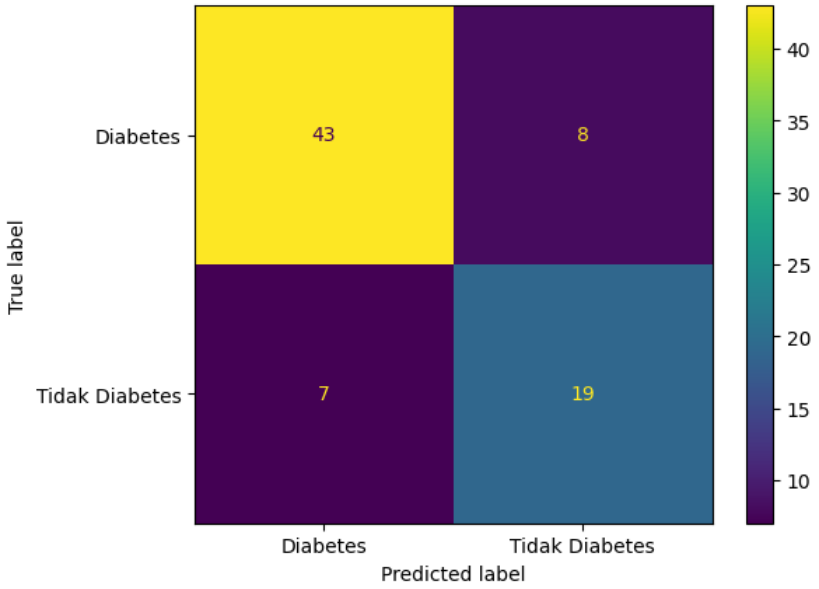
Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Des 17, 2024 | Revised Des 24, 2024 |
Accepted Jan 02, 2025

Hal. 11-20
Vol. 7; No. 1
Februari 2025

Classification Report

	precision	recall	f1-score	support
0.0	0.84	0.87	0.85	107
1.0	0.67	0.62	0.64	47
accuracy			0.79	154
macro avg	0.76	0.74	0.75	154
weighted avg	0.79	0.79	0.79	154

Gambar 6. Hasil Confusion Matrix 80%:20%



Gambar 7. Confusion Matrix 90%:10%

Pada gambar 7. merupakan hasil dari *confusion matrix* dari *naïve bayes* dengan perbandingan data training dan data testing sebesar 90%:10%. Sehingga didapatkan hasil akurasi, presisi, recall dan F1-score sebagai berikut:

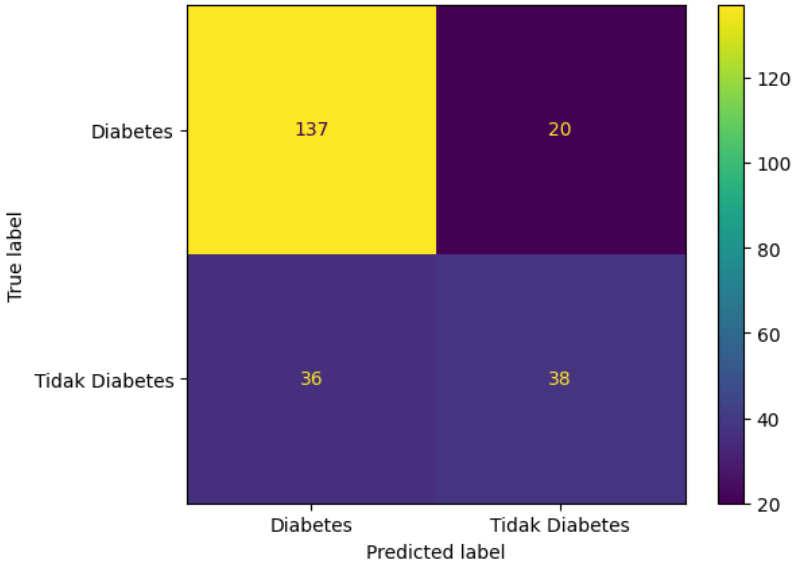
Classification Report

	precision	recall	f1-score	support
0.0	0.86	0.84	0.85	51
1.0	0.70	0.73	0.72	26
accuracy			0.81	77
macro avg	0.78	0.79	0.78	77
weighted avg	0.81	0.81	0.81	77

Gambar 8. Hasil Confusion Matrix 90%:10%

b. Random Forest

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Des 17, 2024 | Revised Des 24, 2024 |
Accepted Jan 02, 2025
Hal. 11-20
Vol. 7; No. 1
Februari 2025



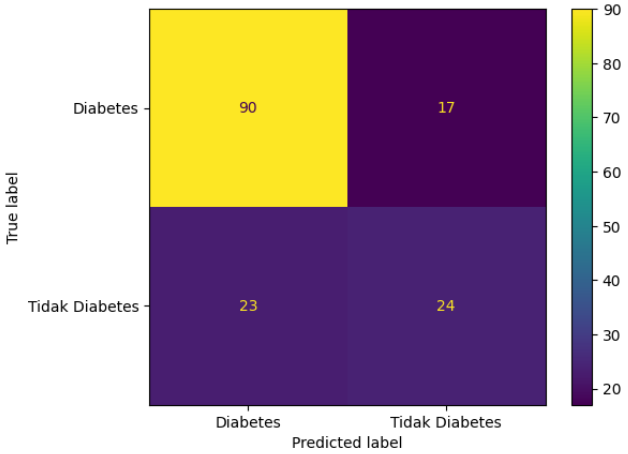
Gambar 9. Hasil *Confusion Matrix Random Forest*

Pada gambar 9 merupakan hasil dari confusion matrix dari random forest dengan perbandingan data training dan data testing sebesar 70%:30%. Sehingga didapatkan hasil akurasi, presisi, recall dan F1-score sebagai berikut:

Classification Report

	precision	recall	f1-score	support
0.0	0.79	0.88	0.83	157
1.0	0.67	0.51	0.58	74
accuracy			0.76	231
macro avg	0.73	0.70	0.71	231
weighted avg	0.75	0.76	0.75	231

Gambar 10. Hasil Confusion Matrix 70%:30%



Gambar 11.Confusion Matrix Random Forest 80%:20%

Pada gambar 11 merupakan hasil dari *confusion matrix* dari *random forest* dengan perbandingan data training dan data testing sebesar 80%:20%. Sehingga didapatkan hasil akurasi, presisi, recall dan F1-score.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian didapatkan metode naïve bayes memiliki rata-rata akurasi tertinggi yaitu 78,66% sedangkan metode random forest hanya 76,33% sehingga dapat disimpulkan bahwa metode naïve bayes merupakan metode terbaik dalam melakukan analisa klasifikasi penyakit diabetes.

Saran Saran atau usulan yang dapat diusulkan pada penelitian ini yaitu pada penelitian selanjutnya dapat membandingkan metode *Random forest* dengan metode yang lain seperti KNN atau metode *Linear Regression* (LR) dan metode *Support Vector Machine* (SVM) untuk membandingkan metode yang lebih baik dari *Random Forest*. Usulan yang lainnya dapat menggunakan studi kasus yang lain dengan jumlah data yang berbeda untuk melihat tingkat akurasinya

DAFTAR PUSTAKA

[1] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes," *Indonesian Journal of Data and Science*, vol. 1, no. 2, hal. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.

[2] W. Apriliah, I. Kurniawan, M. Baydhowi, dan T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *Sistemasi*, vol. 10, no. 1, hal. 163, 2021, doi: 10.32520/stmsi.v10i1.1129.

[3] N. M. Putry, "Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus," *EVOLUSI : Jurnal Sains dan Manajemen*, vol. 10, no. 1, 2022, doi: 10.31294/evolusi.v10i1.12514.

[4] S. P. Nainggolan dan A. Sinaga, "Comparative Analysis of Accuracy of Random Forest and Gradient Boosting Classifier Algorithm for Diabetes Classification," *Sebatik*, vol. 27, no. 1, hal. 97–102, 2023, doi: 10.46984/sebatik.v27i1.2157.

[5] D. Nasien *et al.*, "Perbandingan Implementasi Machine Learning Menggunakan Metode KNN , Naive Bayes , Dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes," vol. 4, no. 1, 2024.

[6] Q. R. Cahyani *et al.*, "Prediksi Risiko Penyakit Diabetes menggunakan Algoritma Regresi Logistik Diabetes Risk Prediction using Logistic Regression Algorithm Article Info ABSTRAK," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 1, no. 2, hal. 2828–9099, 2022, doi: 10.55123/jomlai.v1i2.598.

[7] G. Abdurrahman, "Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier," *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, vol. 7, no. 1, hal. 59–66, 2022, doi: 10.32528/justindo.v7i1.4949.

[8] N. Nurussakinah dan M. Faisal, "Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree," *Jurnal Informatika*, vol. 10, no. 2, hal. 143–149, 2023, doi: 10.31294/inf.v10i2.15989.

[9] R. Supriyadi, W. Gata, N. Maulidah, dan A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis : Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 13, no. 2, hal. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.

[10] A. Prandika Siregar, D. Priyadi Purba, J. Putri Pasaribu, K. Reza Bakara, dan J. V Willem Iskandar Pasar Medan Estate, "Implementasi Algoritma Random Forest Dalam Klasifikasi Diagnosis Penyakit Stroke," *Jurnal Penelitian Rumpun Ilmu Teknik (JUPRIT)*, vol. 2, no. 4, hal. 155–164, 2023, [Daring]. Tersedia pada: https://doi.org/10.55606/juprit.v2i4.3039

[11] N. F. Patimah, M. Abdurrohman, A. R. Rinaldi, dan A. Rinaldi Dikananda, "Implementasi Algoritma Naïve Bayes dalam Klasifikati Penyakit Diabetes," *Jurnal Data Science & Informatika*, vol. 1, no. 1, hal. 6–10, 2021, [Daring]. Tersedia pada: http://publikasi.bigdatascience.id

[12] A. Fauzi dan A. H. Yunial, "Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K – Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 3, hal. 470, 2022, doi: 10.26418/jp.v8i3.56656.

[13] F. Sholekhah, A. D. Putri, dan L. Efrizoni, "Comparison of Naive Bayes and K-Nearest Neighbors Algorithms for Metabolic Syndrome Classification Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbors untuk Klasifikasi Metabolik Sindrom," vol. 4, no. April, hal. 507–514, 2024.

[14] R. Rosita, D. Ananda Agustina Pertiwi, dan O. Gina Khoirunnisa, "Prediction of Hospital Intesive Patients Using Neural Network Algorithm," *Journal of Soft Computing Exploration*, vol. 3, no. 1, hal. 8–11, 2022, doi: 10.52465/joscex.v3i1.61.

[15] S. T. Ratna Patil, "A comparative analysis on the evaluation of classification algorithms in the prediction of diabetes," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 5, hal. 3966–3975, 2018, doi: 10.11591/ijece.v8i5.pp3966-3975.