

PENERAPAN ALGORITMA NAÏVE BAYES MENGGUNAKAN FITUR FORWARD SELECTION PADA PENYAKIT DIABETES

Diterima Redaksi: 23 Desember 2024; Revisi Akhir: 6 Maret 2025; Diterbitkan Online: 8 Juni 2025

Aditya Fauzi Samsuri^{1*)}, Ahmad Faqih²⁾, Kaslani³⁾

^{1, 2,3)} Teknik Informatika, STMIK IKMI Cirebon

³⁾ Komputerisasi Akuntansi, STMIK IKMI Cirebon

^{1, 2,3)} Jalan Perjuangan No. 10B, Kec. Kesambi, Kota Cirebon, Jawa Barat, Indonesia, kode pos: 45135

e-mail: aditfauzi178@gmail.com^{1*)}

Abstrak: Diabetes merupakan salah satu penyakit kronis yang prevalensinya terus meningkat secara global, dan memerlukan diagnosis yang cepat dan akurat untuk mencegah komplikasi lebih lanjut. Dalam upaya meningkatkan akurasi diagnosis, penelitian ini menerapkan metode forward selection dengan algoritma Naive Bayes untuk klasifikasi data diabetes. Fitur forward selection digunakan dalam pemilihan fitur untuk menyaring atribut-atribut yang relevan sehingga dapat meningkatkan performa model. Penelitian ini menggunakan teknik analisis berbasis KDD (Knowledge Discovery in Databases) yang diawali dengan pengumpulan data dari sumber dataset terbuka yang berisi informasi medis pasien diabetes. Data kemudian diolah dengan preprocessing yang meliputi penanganan missing value dan set role. Langkah selanjutnya adalah penerapan fitur forward selection untuk memilih fitur yang paling berpengaruh terhadap prediksi diabetes. Algoritma Naive Bayes kemudian diaplikasikan pada subset fitur yang terpilih. Performa model diukur menggunakan metrik akurasi, precision, dan recall. Hasil penelitian menunjukkan bahwa penggunaan fitur forward selection berhasil meningkatkan akurasi model Naive Bayes secara signifikan dibandingkan dengan model yang menggunakan seluruh fitur tanpa seleksi. Pada model yang diterapkan dengan seleksi fitur, akurasi yang diperoleh mencapai 75.01%, precision 76.10%, dan recall 90.00%, sementara model tanpa seleksi fitur hanya mencapai akurasi 67.32%, precision 63.84 %, dan recall 64.27%. Kesimpulan dari penelitian ini adalah bahwa fitur forward selection pada algoritma Naive Bayes mampu meningkatkan akurasi dalam diagnosis penyakit diabetes dan dapat mengurangi jumlah fitur untuk mengetahui atribut yang berpengaruh dalam penelitian yang dilakukan.

Kata Kunci— Diabetes, Forward Selection, Naive Bayes, Diagnosis Diabetes, Akurasi Model

Abstract: Diabetes is one of the chronic diseases whose prevalence continues to increase globally and requires fast and accurate diagnosis to prevent further complications. To improve diagnostic accuracy, this study applies the forward selection method with the Naive Bayes algorithm for diabetes data classification, the forward selection is used in feature selection to filter relevant attributes so that it can improve model performance. The analysis technique of this study is based on KDD which begins with data collection from open dataset sources containing medical information on diabetes patients. The data is then processed with preprocessing which includes handling missing values and set roles. The next step is the application of the Forward Selection feature to select the features that have the most influence on diabetes prediction. The Naive Bayes algorithm is then applied to the selected feature subset. Model performance is measured using accuracy, precision, and recall metrics. The results show that the use of the Forward Selection feature has succeeded in significantly increasing the accuracy of the Naive Bayes model compared to models that use all features without selection. In the model applied with feature selection, the accuracy obtained reached 75.01%, precision 76.10%, and recall 90.00%, while the model without feature selection only achieved 67.32% accuracy, 63.84% precision, and 64.27% recall. The conclusion of this study is that the application of the forward selection to the Naive Bayes algorithm is able to increase the accuracy in diagnosing diabetes and can reduce computational complexity by using fewer features.

Keywords— Diabetes, Forward Selection, Naive Bayes, Diabetes Diagnosis, Model Accuracy

I. PENDAHULUAN

Dalam beberapa dekade terakhir, perkembangan di bidang informatika telah meningkatkan dampak yang signifikan pada kehidupan manusia, mulai dari teknologi hingga bisnis dan Pendidikan. Inovasi dalam pengolahan data dan analisis algoritma dapat meningkatkan efisiensi dan akurasi di banyak sektor [1]. Salah satu bidang yang mendapat manfaat besar dari kemajuan ini adalah kesehatan, di mana teknologi informasi digunakan untuk mendukung diagnosis dan pengelolaan penyakit. Penyakit diabetes, yang menjadi salah satu tantangan kesehatan global, membutuhkan

pendekatan yang lebih tepat dalam diagnosis agar pengobatan dapat dilakukan secara efektif. Dengan menggunakan algoritma canggih seperti Naïve Bayes, diharapkan akurasi dalam mendeteksi gejala diabetes dapat ditingkatkan, sehingga mendukung penanganan yang lebih baik bagi para penderita [2].

Dalam konteks informatika, salah satu permasalahan yang signifikan adalah tingginya tingkat kesalahan dalam diagnosis penyakit, termasuk diabetes. Meskipun teknologi informasi telah berkembang pesat, masih terdapat tantangan dalam menerapkan algoritma yang dapat secara akurat menganalisis gejala dan data medis. Kesenjangan dalam literatur menunjukkan bahwa banyak penelitian yang ada belum mengoptimalkan penggunaan algoritma berbasis machine learning, seperti Naïve Bayes, dalam konteks diagnosis medis [3]. Hal ini menjadi berbahaya mengingat semakin meningkatnya jumlah kasus diabetes di seluruh dunia, yang memerlukan penanganan lebih cepat dan efektif. Di sisi lain, penggunaan metode tradisional dalam diagnosis sering kali menyebabkan keterlambatan dalam pengobatan, yang dapat berdampak serius bagi pasien. Selain itu, ada kebutuhan untuk menyempurnakan model prediksi yang dapat menangani data besar dan beragam, yang sering kali berasal dari sumber yang tidak terstruktur [4].

Dengan memanfaatkan algoritma *Naïve Bayes*, yang dikenal karena efisiensinya dalam mengelola data berukuran besar, kita dapat meningkatkan akurasi diagnosis serta memberikan dukungan yang lebih baik bagi tenaga medis. Mengingat tren global menuju digitalisasi dalam kesehatan, penting untuk mengatasi isu ini agar sistem kesehatan dapat beradaptasi dan berkembang. Upaya untuk mengurangi kesalahan diagnosis melalui penerapan teknologi yang tepat menjadi relevan dan sangat bernilai dalam konteks peningkatan kualitas hidup pasien. Solusi yang inovatif diharapkan dapat mengurangi beban sistem kesehatan dan mempercepat proses pengobatan yang efektif.

Penelitian sebelumnya oleh [5] membahas pengoptimalan klasifikasi untuk prediksi diabetes menggunakan Particle Swarm Optimization (PSO). Studi tersebut mengevaluasi beberapa algoritma, termasuk *Naïve Bayes*, *Random Forest*, *Decision Tree*, dan *K-Nearest Neighbor*, pada dataset yang terdiri dari 520 contoh dan 17 atribut. Hasilnya menunjukkan bahwa algoritma-algoritma tersebut mengalami peningkatan akurasi setelah proses pengoptimalan, dengan *Random Forest* mencapai akurasi tertinggi sebesar 98,35%. Penelitian ini menyoroti pentingnya model klasifikasi yang efektif dalam mendeteksi diabetes secara dini dan merekomendasikan eksplorasi lebih lanjut terhadap metode pengoptimalan.

Penelitian oleh [6] membahas pengembangan dan evaluasi aplikasi berbasis web untuk mendiagnosis hipertensi menggunakan algoritma Naïve Bayes dengan koreksi Laplace. Aplikasi ini mencapai akurasi diagnostik sebesar 77% berdasarkan 30 kasus uji dan menerima skor evaluasi pengguna sebesar 3,85 dari 5. Proses pengembangan aplikasi mengikuti model *Software Development Life Cycle* (SDLC) dengan pendekatan *Waterfall* yang mencakup pengumpulan persyaratan, desain, pengodean, pengujian, dan penerapan. Aplikasi ini dirancang untuk meningkatkan kesadaran masyarakat tentang hipertensi sekaligus menyediakan sumber daya edukasi dan fungsi diagnosis.

Penelitian lain oleh [7] menggunakan teknik klasifikasi untuk mendiagnosis penyakit scabies pada hewan peliharaan dengan algoritma Naïve Bayes. Proses penelitian mengikuti pendekatan Knowledge Discovery in Database (KDD), meliputi pemilihan data, pembersihan, pembagian, penambangan, dan evaluasi model. Penelitian menghasilkan *accuracy* sebesar 97,20%, *precision* 88,00%, dan *recall* 93,62%. Dari total 1.208 data yang diuji, 912 hewan terdiagnosis positif dan 157 negatif, dengan gejala telinga keropeng sebagai indikator utama untuk diagnosis positif. Studi ini juga merekomendasikan pemeriksaan lebih lanjut oleh dokter hewan untuk hasil yang lebih akurat.

Penelitian ini bertujuan meningkatkan akurasi diagnosis diabetes dengan menerapkan algoritma Naïve Bayes untuk menganalisis gejala yang relevan. Dengan menggunakan teknik *machine learning* pada data sekunder yang diproses melalui software RapidMiner, penelitian ini mengevaluasi akurasi model melalui metrik seperti *accuracy*, *precision*, dan *recall*, serta mengidentifikasi gejala yang signifikan dalam diagnosis. RapidMiner digunakan untuk mempermudah proses pengolahan data, pemilihan fitur, dan penerapan algoritma, sehingga menghasilkan analisis yang lebih cepat dan akurat. Hasilnya diharapkan dapat memberikan kontribusi pada pengembangan sistem diagnosis berbasis teknologi yang lebih efektif, mendukung tenaga medis dalam pengambilan keputusan, dan meningkatkan kualitas perawatan pasien dengan mengurangi kesalahan diagnostik.

II. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah proses untuk menggali atau mengumpulkan informasi penting dari data besar (Big Data). Proses ini biasanya menggunakan metode matematika, statistik, dan bahkan kecerdasan buatan (AI). Beberapa teknik yang digunakan dalam data mining meliputi klasifikasi, pengelompokan, asosiasi, regresi, peramalan, analisis urutan, dan analisis deviasi. Data mining dilakukan dengan menerapkan teknik dan algoritma tertentu untuk mengekstraksi informasi yang bernilai dari kumpulan data yang ada. Tujuan utama dari proses ini adalah untuk menemukan informasi yang tersembunyi dalam data, yang dapat dimanfaatkan di berbagai bidang, seperti keuangan, bisnis, dan kesehatan.

2.2. Naïve Bayes

Algoritma Naive Bayes Classifier adalah salah satu metode yang digunakan dalam teknik klasifikasi. Naive Bayes adalah metode pengklasifikasian yang mengandalkan probabilitas dan statistik, yang pertama kali diajukan oleh ilmuwan asal Inggris, Thomas Bayes. Algoritma ini berfungsi untuk memprediksi kemungkinan di masa depan berdasarkan data atau pengalaman yang ada sebelumnya, sehingga dikenal dengan nama Teorema Bayes [8][9].

Rumus Utama Teorema Bayes:

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

Keterangan:

$P(H|X)P(H|X)P(H|X)$ = Probabilitas hipotesis **H** (kelas) berdasarkan bukti **X** (fitur).

$P(X|H)P(X|H)P(X|H)$ = Probabilitas munculnya bukti **X** jika hipotesis **H** benar.

$P(H)P(H)P(H)$ = Probabilitas awal dari hipotesis **H** (prior probability).

$P(X)P(X)P(X)$ = Probabilitas total dari bukti **X** (dapat diabaikan dalam klasifikasi karena tetap).

2.3. Rapid Miner

RapidMiner adalah sebuah platform perangkat lunak untuk analisis data yang dikembangkan oleh perusahaan dengan nama yang sama. Platform ini menyediakan lingkungan yang terpadu untuk mempersiapkan data, melakukan pembelajaran mesin, menerapkan pembelajaran mendalam, menambang teks, dan melakukan analisis prediktif [10].

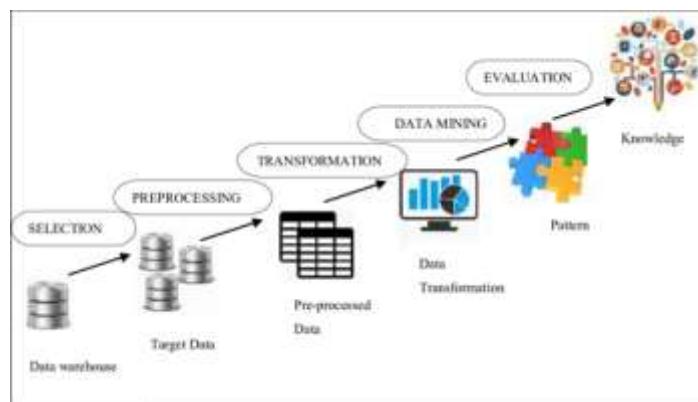
2.4. Forward Selection

Forward Selection adalah salah satu metode dalam feature selection yang bertujuan untuk memilih subset fitur terbaik dari sekumpulan fitur yang tersedia. Metode ini bekerja dengan menambahkan fitur satu per satu berdasarkan kontribusinya terhadap performa model [11].

III. METODE PENELITIAN

3.1. Metode Penelitian

Metode yang digunakan penelitian ini adalah eksperimental. Dengan menggunakan strategi sains untuk menguji teori dan mengidentifikasi hubungan kausal antara variabel tertentu. Dengan menggunakan variabel Pregnancies, Glucose, Blood Pressure, Insulin, BMI, Diabetes Pedigree Function, Age, dan Outcome sebagai variabel label, data dari situs web dataset Diabetes kaggle akan diproses menggunakan algoritma Naive Bayes untuk memprediksi tingkat akurasi diabetes [11].



Gambar 1. Tahapan Metode Penelitian

Gambar 1 menampilkan alur tahapan dalam proses KDD (Knowledge Discovery in Databases) dengan menggunakan software RapidMiner dengan menggunakan fitur seperti *Read CSV*, *Replace Missing Values*, *Set Role*, *Multiply*, *Naïve Bayes*, *Cross Validation*, *Forward Selection* untuk mempermudah implementasi. Berikut adalah cara setiap tahap dilakukan:

A. *Import Data*

1. Menggunakan operator *Read CSV* dataset dapat diimpor langsung ke RapidMiner.
2. Format data yang didukung mencakup CSV, Excel, database, atau file teks lainnya.

B. *Preprocessing Data*

1. *Replace Missing Values*: Operator ini digunakan untuk mengisi nilai yang hilang dengan metode seperti rata-rata, median, atau mode.
2. *Set Role*: Digunakan untuk menentukan atribut target (label) dan atribut fitur.

C. *Modeling*

1. *Naïve Bayes*: Operator ini diterapkan untuk membuat model klasifikasi.
2. *Multiply*: Digunakan jika data atau model perlu direplikasi untuk analisis lebih lanjut.
3. *Forward Selection*: Untuk memilih atribut yang paling relevan dan meningkatkan kinerja model.

D. *Performance Evaluation*

1. *Cross Validation*: Operator ini membagi data menjadi beberapa lipatan (fold) untuk melatih dan menguji model secara bergantian, memastikan hasil yang lebih andal.
2. *Performance (Classification)*: Menyediakan metrik evaluasi seperti akurasi, precision, recall, untuk menilai kinerja model.

3.2. Sumber Data

Sumber data penelitian ini berasal dari dataset Kaggle. Atribut yang digunakan sebanyak 8 atribut yang terdiri dari Pregnancies, Glucose, Blood Pressure, Insulin, BMI, Age, Diabetes Pedigree Function, dan Outcome sebagai label dengan jumlah data sebanyak 768. Kumpulan data ini yang diambil dari populasi pasien dengan kondisi atau faktor risiko yang dapat berhubungan dengan diabetes.

Tabel 1. Atribut Sample Data

No	Pregnancies	Glucose	BloodPressure	Insulin	BMI	Age	Diabetes Pedigree Function	Outcome
1	6	148	72	0	33.06.00	50	0,435417	Positive
2	1	85	66	0	26.06.00	31	0,24375	Negative
3	8	183	64	0	23.03	32	0,466667	Positive
...
766	5	121	72	122	26.02.00	30	0,170139	Negative
767	1	126	60	0	30.01.00	47	0,242361	Positive
768	1	93	70	0	30.04.00	23	0,21875	Negative

3.2. Teknik Analisis data

Klasifikasi dengan menggunakan algoritma Naïve Bayes adalah salah satu metode klasifikasi yang sering digunakan dalam Data Mining. Naïve Bayes merupakan prosedur statistik untuk klasifikasi yang memungkinkan untuk mengungkapkan ketidakpastian terkait suatu model, berdasarkan penjelasan hasil probabilitas. Metode ini sangat berguna untuk menyelesaikan masalah diagnosis serta peramalan atau prediksi.

IV. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1. Data Selection

Langkah utama untuk mengelola data dengan menggunakan port Read CSV. Kemudian dataset dimasukkan dengan cara mengklik bagian Read CSV, yang kemudian akan tampil lokasi data yang akan dimasukkan. Selanjutnya, gunakan yang akan diolah sebelumnya dengan tools Read CSV. Penting untuk mengevaluasi data yang akan digunakan terlebih dahulu, kemudian melanjutkan dengan proses pelabelan data. Atribut yang diberikan label dalam penelitian ini adalah Outcome, yang menunjukkan apakah hasilnya negatif atau positif. Jika terkena diabetes akan diberi label "Positive" sedangkan yang tidak akan diberi label "Negatif". Gambar 2 di bawah ini Read CSV merupakan tahapan dan data yang digunakan untuk proses pada model Import Data.



Gambar 2. Read CSV

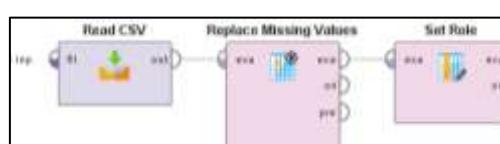
Row No.	Outcome	Pregnancies	Glucose	BloodPress...	Insulin	BMI	DiabetesPe...	Age
1	Positive	6	148	72	0	33.80	0.435410887	39
2	Negative	1	89	66	0	28.80	0.24578	31
3	Positive	8	183	64	0	23.03	0.488888887	32
4	Negative	1	88	66	94	38.01	0.118873022	21
5	Positive	0	127	40	168	33.01	2.268	33
6	Negative	5	116	74	0	25.06	0.136883333	30
7	Positive	3	78	50	58	31	0.172222222	25
8	Negative	18	118	8	0	38.03	0.0939393939	29
9	Positive	2	197	70	543	36.06	0.109722222	33
10	Positive	0	128	64	0	0	0.161111111	54
11	Negative	6	112	92	0	37.06	0.132638889	30
12	Positive	10	188	74	0	38	0.272818667	34
13	Unknown	11	186	85	0	37.31	0.111111111	37

Gambar 3. Sample Data Read CSV

Read CSV di RapidMiner berfungsi untuk mengetahui isi dari dataset berformat file. Read CSV sering digunakan karena mudah diproses oleh banyak aplikasi dan digunakan secara luas untuk menyimpan dataset dalam format yang sederhana dan ringan.

4.1.2. Preprocessing

Langkah kedua adalah melakukan Replace Missing Values ditambahkan untuk menghapus data kosong pada isi dataset. Selain itu, penggunaan fitur Set Role akan membantu dalam mengidentifikasi dan memisahkan tujuan penggunaan setiap fitur dalam model. Berikut adalah gambar dari sub proses. Gambar 3 merupakan penjelasan tentang Replace Missing Values dan Set Role merupakan tahapan proses pada model Preprocessing Data.



Gambar 3. Proses Preprocessing

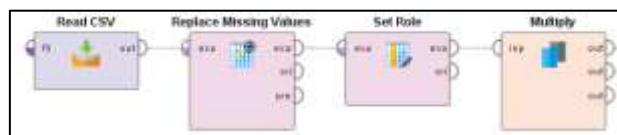
Name	Type	Missing	Statistics	Filter (3 / 8 attributes)	General Statistics
Outcome	Normal	0	Positive (268)	Positive (268)	Positive
Pregnancies	Integer	0	0-9	0-9	0-9
Glucose	Integer	0	0-100	0-100	0-100
BloodPressure	Integer	0	0-99	0-99	0-99
Insulin	Integer	0	0-846	0-846	0-846
BMI	Normal	0	07.01-67.11	07.01-67.11	07.01-67.11
DiabetesPedigreeFunction	Normal	0	2.409 (1)	0.176388889 (6)	0.176388889

Gambar 4. Hasil Data Proses Preprocessing

Berdasarkan hasil preprocessing data yang ditampilkan, tidak terdapat data yang kosong dalam dataset tersebut. Semua atribut memiliki nilai yang terisi secara lengkap. Selain itu, atribut Outcome telah dipilih sebagai label atau target dalam analisis ini, di mana nilai Positive (268) dan Negative (500) menunjukkan kategori klasifikasi yang akan digunakan dalam pemodelan atau analisis lebih lanjut.

4.1.3. Transformation

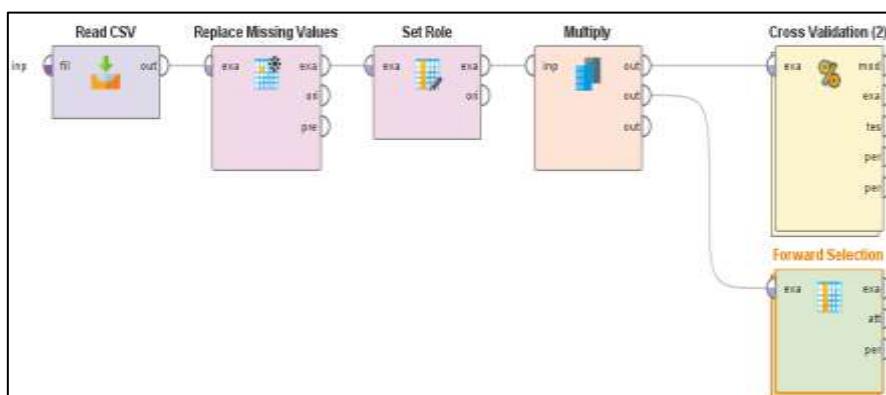
Tahap selanjutnya adalah transformation, Multiply digunakan untuk menduplikasi data atau membuat beberapa salinan dataset yang akan dioperasikan dengan menggunakan Naïve Bayes dan Naïve Bayes + Forward Selection. Operator ini sering digunakan ketika ada kebutuhan untuk menjalankan beberapa alur pemrosesan yang berbeda pada dataset yang sama tanpa harus mengimport ulang data atau menjalankan proses dari awal. Gambar 4 dibawah ini Multiply termasuk dalam tahapan proses Modeling.



Gambar 5. Proses Transformation

4.1.4. Data Mining

Langkah keempat adalah menggunakan Cross Validation untuk menguji performa model Naive Bayes secara objektif dan mengukur kemampuannya dalam generalisasi terhadap data baru. Proses ini dilakukan dengan membagi dataset ke dalam beberapa subset (10 folds) dan melatih serta menguji model pada setiap fold. Forward Selection diterapkan untuk memilih fitur-fitur yang paling berkontribusi terhadap performa model, sehingga model menjadi lebih sederhana dan akurat. Pemilihan fitur ini dilakukan dengan menambahkan fitur satu per satu ke dalam model, lalu mengukur peningkatan performa setelah penambahan setiap fitur. Gambar 5 dibawah ini Forward Selection termasuk dalam tahapan proses Modeling.



Gambar 6. Permodelan Data Mining

4.1.5. Data Evaluasi

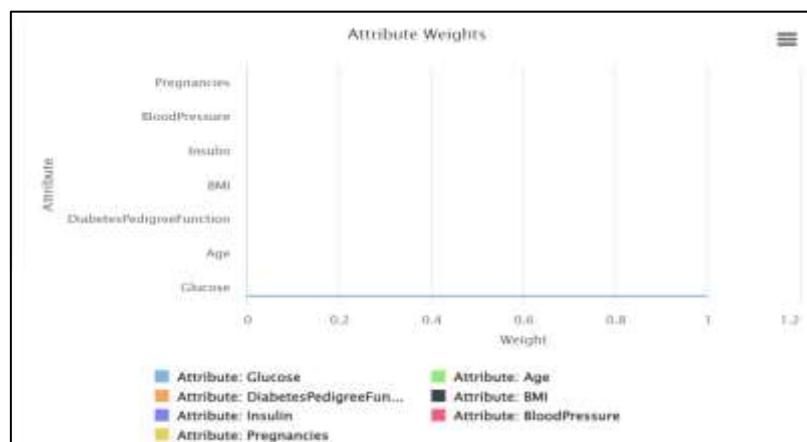
Hasil yang telah dilakukan pengujian model Naive Bayes dan Forward Selection dengan metode Cross Validation untuk memastikan model stabil dan dapat diandalkan dalam membuat prediksi. Nilai evaluasi berikut merupakan hasil perbandingan dari penerapan Naïve Bayes & Naïve Bayes + Forward Selection. Tabel 2 dibawah ini termasuk tahapan Performance Evaluation.

Tabel 2. Hasil perbandingan antara Naïve Bayes & Naïve Bayes + Forward Selection

Model	Accuracy	Precision	Recall
Naïve Bayes	67.32%	74.89%	74.42%
Naïve Bayes + Forward Selection	75.01%	76.10%	90.00%

Tabel 2 menjelaskan bagaimana hasil penggunaan Forward Selection dalam pemilihan fitur memberikan dampak positif pada performa model Naïve Bayes, dengan peningkatan signifikan pada semua metrik evaluasi, *Accuracy* yang naik dari 67.27% menjadi 75.01%, *Precision* naik dari 63.27% menjadi 76.10%, *Recall* naik dari 64.27% menjadi 90.00% Hal ini menunjukkan bahwa Forward Selection membantu model memfokuskan pada fitur-fitur yang paling relevan, sehingga meningkatkan kualitas prediksi dan mengurangi gangguan dari fitur yang kurang penting.

Berdasarkan teknik Forward Selection, fitur-fitur yang memiliki kontribusi signifikan terhadap performa model dipilih untuk menyederhanakan model dan mempertahankan akurasinya. Proses ini menunjukkan bahwa fitur-fitur tertentu lebih relevan untuk meningkatkan kinerja model. Gambar 6 di bawah ini termasuk tahapan Performance Evaluation.



Gambar 6. Attribute Berpengaruh

Gambar 6 menjelaskan bahwa dari beberapa atribut yang di gunakan seperti Pregnancies, BloodPressure, Insulin, BMI, DiabetesPedigreeFunction, Age dan atribut Glucose menunjukkan kontribusi yang paling besar dalam memprediksi diabetes dibandingkan atribut lainnya. Hal ini dikarenakan Glucose memiliki hubungan langsung dan signifikan terhadap risiko diabetes, sementara atribut lainnya hanya berperan sebagai faktor pendukung dengan kontribusi yang lebih kecil.

4.2. Pembahasan

4.2.1. Pengukuran Akurasi Menggunakan Confusion Matrix

Confusion Matrix adalah metode untuk mengevaluasi kinerja klasifikasi, di mana hasil klasifikasi dipengaruhi oleh tingkat akurasi. Metode ini didasarkan pada perbandingan antara hasil klasifikasi yang dihasilkan oleh model dengan hasil klasifikasi yang sesungguhnya. Confusion Matrix memberikan informasi mengenai perbandingan tersebut dan sangat penting karena dapat menunjukkan seberapa baik performa model berdasarkan pengukuran akurasi yang telah dilakukan sebelumnya yang mendapatkan *Accuracy* 67.32% dan 75.01%, *Precision* 74.89% dan 76.10%, dan *Recall* 74.42% dan

90.00%. Kinerja model klasifikasi pada data uji dengan nilai yang sudah diketahui dapat digambarkan menggunakan matriks ini, yang digunakan untuk menghitung tingkat ketepatan model.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Gambar 7. Confusion Matrix

Rumus yang digunakan untuk menghitung *accuracy*:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Rumus yang digunakan untuk menghitung *precision*:

$$Precision = \frac{TP}{TP + TN}$$

Rumus yang digunakan untuk menghitung *recall*:

$$Recall = \frac{TP}{TP + FN}$$

V. KESIMPULAN DAN SARAN

Kesimpulan yang dapat diambil berdasarkan dalam penelitian ini yaitu berdasarkan hasil penelitian ini menunjukkan bahwa menggunakan Forward Selection untuk pemilihan fitur dalam model Naive Bayes dapat memberikan peningkatan akurasi sebesar 75.01% , precision 76.10% , dan recall 90.00% dibandingkan tanpa menggunakan fitur tambahan yang hanya menghasilkan akurasi . Teknik ini membantu dalam meningkatkan performa model, sesuai dengan temuan dalam literatur terdahulu.

Untuk penelitian selanjutnya, disarankan agar peneliti mencoba algoritma lain selain Naive Bayes agar dapat mengetahui akurasi dan presisi yang beragam. Hal ini memungkinkan perbandingan yang lebih menyeluruh terhadap performa algoritma dalam menentukan kualitas prediksi.

DAFTAR PUSTAKA

- [1] M. F. Rizqullah, N. T. Raihana, and M. I. Jambak, “Komparasi Penerapan Algoritma C4.5, K-Nearest Neighbor, dan Naïve Bayes Untuk Keberlangsungan Pasien Gagal Jantung,” *Klik Kajian Ilmiah. Informatika. Dan Komputer.*, vol. 4, no. 5, pp. 2580–2587, 2024, doi: 10.30865/klik.v4i5.1788.
- [2] F. Novaldy and A. Herliana, “Penerapan Pso Pada Naïve Bayes Untuk Prediksi Harapan Hidup Pasien Gagal Jantung,” *Jurnal. Responsif Riset. Sains dan Informatika.*, vol. 3, no. 1, pp. 37–43, 2021, doi: 10.51977/jti.v3i1.396.
- [3] M. Yusa, F. F. Coastera, and M. R. Yandika, “Reduksi Dimensi Data menggunakan Metode Wrapper Sequential Feature Selection untuk Peningkatan Performa Algoritma Naïve Bayes terhadap Dataset Medis,” *Jurnal. Edukasi dan Penelitian. Informatika.*, vol. 8, no. 2, pp. 364–370, 2022, doi: 10.26418/jp.v8i2.54328.
- [4] H. Fahmi and Sutisna, “Implementasi Data Mining Klasifikasi Gejala Penyakit TB Menggunakan Algoritma Naive Bayes pada Studi Kasus,” *Jurnal. Indonesia. Manajemen. Informatika. dan Komunikasi.*, vol. 5, no. 3, pp. 2888–2898, 2024.
- [5] A. Fauzi and 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.*, vol. 8, no. 3, pp. 470–481, 2022, doi: 10.26418/jp.v8i3.56656.

- [6] J. Lemantara, “Rancang Bangun Aplikasi Hipertensi. EDU Sebagai Media Edukasi Dan Diagnosis Penyakit Hipertensi Menggunakan Metode Naïve Bayes Dengan Laplace Correction,” *INFOTECH Jurnal. Informatika. Teknologi.*, vol. 5, no. 1, pp. 146–160, 2024, doi: 10.37373/infotech.v5i1.1197.
- [7] S. Marwah, R. Astuti, and F. M. Basysyar, “Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Diagnosis Penyakit Kulit Scabies Pada Hewan,” *JATI (Jurnal Mahasiswa. Teknik. Informatika.)*, vol. 7, no. 6, pp. 3892–3897, 2023, doi: 10.36040/jati.v7i6.8276.
- [8] H. Nugroho, G. E. Yuliastuti, and A. Firman, “Klasifikasi Diagnosis Diabetes Melitus Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Backward Elimination,” *Jurnal. Ilmiah. NERO*, vol. 8, no. 2, p. 2023, 2023, [Online]. Available: <https://journal.trunojoyo.ac.id/nero/article/view/21110>
- [9] A. Nur Kirana, B. Nurhakim, S. Eka Permana, W. Prihartono, and G. Dwilestari, “Implementasi Algoritma Naive Bayes Untuk Memprediksi Cuaca Menggunakan Rapidminer,” *JATI (Jurnal Mahasiswa. Teknik. Informatika.)*, vol. 8, no. 2, pp. 1637–1642, 2024, doi: 10.36040/jati.v8i2.8967.
- [10] P. D. Rinanda, B. Delvika, S. Nurhidayarnis, N. Abror, and A. Hidayat, “Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes Pada Ibu Hamil,” *MALCOM Indonesia. Jurnal. Machine. Learning. Computer. Sci.*, vol. 2, no. 2, pp. 68–75, 2022, doi: 10.57152/malcom.v2i2.432.
- [11] M. F. Nugroho and S. Wibowo, “Fitur Seleksi Forward Selection Untuk Menetukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes,” *Jurnal. Informatika. Upgris*, vol. 3, no. 1, pp. 63–70, 2017, doi: 10.26877/jiu.v3i1.1669.
- [12] C. F. Ahmad, N. Suarna, and G. Dwilestari, “Klasifikasi Data Kemiskinan Menggunakan Metode Naïve Bayes Untuk Mengetahui Tingkat Kemiskinan Studi Kasus: Desa Karangasem Kecamatan Leuwimunding Majalengka,” *Jurnal. Informatika. dan Teknologi. Informasi.*, vol. 2, no. 2, pp. 203–208, 2023, doi: 10.56854/jt.v2i2.190.
- [13] A. Byna and M. Basit, “Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naive Bayes,” *Jurnal. Sisfokom (Sistem Informasi. dan Komputer)*, vol. 9, no. 3, pp. 407–411, 2020, doi: 10.32736/sisfokom.v9i3.1023.
- [14] M. D. NurmalaSari, K. Kusrini, and S. Sudarmawan, “Komparasi Algoritma Naive Bayes dan K-Nearest Neighbor Untuk Membangun Pengetahuan Diagnosa Penyakit Diabetes,” *Jurnal. Komtika (Komputasi dan Informatika.)*, vol. 5, no. 1, pp. 52–59, 2021, doi: 10.31603/komtika.v5i1.5140.
- [15] N. M. Putry and B. N. Sari, “Komparasi Algoritma Knna Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus,” *EVOLUSI Jurnal. Sains dan Manajemen.*, vol. 10, no. 1, pp. 45–57, 2022, doi: 10.31294/evolusi.v10i1.12514.
- [16] A. Riani, Y. Susianto, and N. Rahman, “Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes,” *Jurnal. Innovation. Information. Technology. Application.*, vol. 1, no. 1, pp. 25–34, 2019, doi: 10.35970/jinita.v1i01.64.
- [17] A. Ridwan, “Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus,” *Jurnal. SISKOM-KB (Sistem Komputer. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.