

Optimasi Integrasi Adaptif Metode Naive Bayes dan Information Gain untuk Prediksi Komplikasi Diabetes

Adaptive Integration Optimization of Naïve Bayes and Information Gain for Diabetes Complication Prediction

¹Farid Fitriyadi*, ²Indriyati

¹Program Studi Informatika, Universitas Sahid Surakarta

²Program Studi Ilmu Keperawatan, Universitas Sahid Surakarta

^{1,2}Jl. Adi Sucipto No.154, Jajar, Kec. Laweyan, Kota Surakarta, Jawa Tengah 57144

*e-mail: faridfitriyadi@gmail.com

(received: 3 September 2025, revised: 9 September 2025, accepted: 10 September 2025)

Abstrak

Diabetes melitus merupakan masalah kesehatan global dengan prevalensi yang terus meningkat di Indonesia, termasuk di wilayah Surakarta. Komplikasi diabetes menjadi penyebab utama kematian dan penurunan kualitas hidup pasien, sekaligus menimbulkan beban ekonomi yang signifikan. Kondisi ini menuntut adanya model prediksi yang akurat untuk mengidentifikasi pasien berisiko tinggi maupun rendah, terutama di fasilitas kesehatan primer yang masih menghadapi keterbatasan dalam melakukan prediksi secara komprehensif. Penelitian ini bertujuan mengembangkan model prediksi komplikasi diabetes dengan mengoptimalkan integrasi adaptif metode Naive Bayes dan Information Gain. Data yang digunakan mencakup variabel demografis (usia, jenis kelamin), klinis (kadar gula darah, HbA1c, tekanan darah), riwayat kesehatan (keluarga diabetes), dan gaya hidup (aktivitas fisik, pola makan). Hasil implementasi menunjukkan bahwa algoritma Naive Bayes murni menghasilkan akurasi sebesar 75%. Setelah dilakukan seleksi fitur menggunakan Information Gain, akurasi meningkat menjadi 87,5%. Sehingga terjadi peningkatan 12,5%. Temuan ini membuktikan bahwa integrasi Naive Bayes dan Information Gain mampu menghasilkan model prediksi komplikasi diabetes yang lebih akurat, serta berpotensi menjadi alat bantu efektif bagi tenaga kesehatan dalam pencegahan komplikasi dan pengurangan beban sistem kesehatan.

Kata kunci: prediksi, komplikasi diabetes, optimasi, naïve bayes, information gain

Abstract

Diabetes mellitus is a major global health issue with a steadily increasing prevalence in Indonesia, including the Surakarta region. Diabetes complications are among the leading causes of mortality and reduced quality of life in patients, while also imposing a significant economic burden. This situation highlights the need for an accurate predictive model to identify both high- and low-risk patients, particularly in primary healthcare facilities that still face limitations in conducting comprehensive risk prediction. This study aims to develop a predictive model for diabetes complications by optimizing the adaptive integration of the Naïve Bayes algorithm with Information Gain for feature selection. The dataset used includes demographic variables (age, gender), clinical data (blood glucose level, HbA1c, blood pressure), medical history (family history of diabetes), and lifestyle factors (physical activity, dietary patterns). The results indicate that the pure Naïve Bayes algorithm achieved an accuracy of 75%. After applying Information Gain for feature selection, the accuracy improved to 87.5%, representing a 12.5% increase. These findings demonstrate that integrating Naïve Bayes with Information Gain can produce a more accurate prediction model for diabetes complications, making it a potentially effective decision-support tool for healthcare professionals in preventing complications and reducing the burden on the healthcare system.

Keywords: prediction, diabetes complications, optimization, naïve bayes, information gain

1 Pendahuluan

Diabetes melitus merupakan tantangan kesehatan global dengan prevalensi yang terus meningkat. Di Indonesia, jumlah penderita diabetes mencapai 10,7 juta pada tahun 2023 dan diproyeksikan meningkat menjadi 13,7 juta pada tahun 2030 [1]. Di wilayah Surakarta, prevalensi mencapai 2,1% dan menunjukkan tren kenaikan [2]. Dampak lanjut penyakit ini berkontribusi signifikan terhadap morbiditas dan mortalitas. Studi epidemiologi mencatat bahwa sekitar 50% pasien mengalami gangguan mikrovaskular dan makrovaskular dalam kurun waktu 10 tahun setelah diagnosis [3]. Di RSUD Dr. Moewardi, sebanyak 60% pasien diabetes yang dirawat inap telah mengalami komplikasi dengan tingkat kematian mencapai 15% [4]. Komplikasi diabetes menjadi penyebab utama morbiditas dan mortalitas pasien. Data epidemiologi menunjukkan bahwa sekitar 50% pasien diabetes mengembangkan komplikasi setelah 10 tahun terdiagnosis [12]-[13].

Pencegahan risiko melalui deteksi dini merupakan strategi kunci dalam pencegahan dan intervensi tepat waktu [5]. Namun, fasilitas kesehatan primer menghadapi tantangan dalam memprediksi secara akurat pasien berisiko tinggi akibat keterbatasan sistem informasi dan penilaian klinis yang masih subjektif [6]. Penelitian ini mengusulkan pengembangan model prediksi komplikasi diabetes menggunakan integrasi adaptif metode Naive Bayes dan Information Gain [7]. Pendekatan ini berpotensi meningkatkan akurasi prediksi dengan memanfaatkan kelebihan kedua metode dalam klasifikasi data kesehatan [8].

Permasalahan terjadi pada kebutuhan mendesak untuk meningkatkan kapasitas fasilitas kesehatan primer dalam mendeteksi dini dan mencegah komplikasi diabetes [9]-[10]. Model prediksi yang dihasilkan diharapkan dapat menjadi sistem pendukung keputusan klinis yang membantu tenaga kesehatan di Klinik Pratama untuk mengidentifikasi pasien berisiko tinggi, sehingga intervensi pencegahan dapat dilakukan lebih dini dan terarah. Hal ini berpotensi menurunkan angka komplikasi, rawat inap, kecacatan, dan kematian akibat diabetes, serta mengurangi beban ekonomi pada sistem kesehatan [10]-[11]. Penelitian ini juga sejalan dengan prioritas nasional dalam pengendalian penyakit tidak menular dan penguatan sistem kesehatan primer.

Rumusan masalah penelitian ini mencakup: (1) ketiadaan model prediksi risiko yang sederhana, akurat, dan mudah diterapkan di layanan kesehatan primer; (2) pemanfaatan teknik seleksi fitur untuk mengekstraksi variabel paling relevan masih terbatas serta (3) kebutuhan sistem pendukung keputusan yang dapat mengurangi ketergantungan pada penilaian subjektif tenaga kesehatan [12].

Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi komplikasi diabetes dengan mengintegrasikan metode Naive Bayes dan Information Gain secara adaptif, sehingga menghasilkan klasifikasi risiko yang lebih akurat [7], [8]. Manfaat penelitian ini adalah: (i) menghadirkan pendekatan komputasi yang ringan namun efektif untuk fasilitas kesehatan dengan sumber daya terbatas [13]-[14]; (ii) menghasilkan variabel kunci (demografis, klinis, riwayat kesehatan, dan gaya hidup) yang dapat digunakan tenaga medis sebagai dasar intervensi dini [5], [8]; serta (iii) mendukung prioritas nasional dalam pengendalian penyakit tidak menular melalui penguatan layanan kesehatan primer [15].

2 Tinjauan Literatur

Penelitian ini memiliki keterkaitan dengan penelitian terdahulu seperti penelitian dengan judul, *"Predicting Type 2 diabetes onset age using machine learning: A case study in KSA"*, Tujuan: Untuk memprediksi usia onset diabetes tipe 2 di Arab Saudi untuk mendukung intervensi dini. Metode: MLR, ANN, RF, SVR, dan DTR. Hasil: Menunjukkan MLR dan RF memiliki akurasi terbaik, dengan faktor utama yang mempengaruhi usia onset adalah trigliserida, kolesterol, HDL, ferritin, BMI, dan tekanan darah. Persamaan: penggunaan pembelajaran mesin. Perbedaan: Terletak pada fokus pada usia onset diabetes, bukan komplikasi [16].

Penelitian terdahulu dengan judul: *"A diabetes prediction model based on Boruta feature selection and ensemble learning"* Tujuan: Untuk meningkatkan akurasi deteksi dini diabetes dengan teknik Boruta feature selection dan ensemble learning. Metode: Mencakup Boruta, K-Means++, dan stacking ensemble. Hasil: menunjukkan akurasi mencapai 98%. Persamaan: Penggunaan pembelajaran mesin, Perbedaan: Pendekatan yang lebih kompleks dengan ensemble, fokus pada deteksi dini diabetes, bukan komplikasi [17].

Penelitian terdahulu dengan judul: “*Advancing diabetes prediction with a progressive self-transfer learning framework for discrete time series data*”. Tujuan: Untuk mengembangkan framework self-transfer learning progresif untuk prediksi diabetes dengan data time series. Metode: BRITS untuk imputasi, LASSO untuk seleksi fitur, dan transfer learning. Hasil: Adanya peningkatan pada AUC dan F1 score. Persamaan: Penggunaan algoritma prediktif untuk diabetes. Perbedaan: Terdapat pada penggunaan data time series dan deep learning yang kompleks[18]. Penelitian dengan judul : “*A Combined Method for Diabetes Mellitus Diagnosis Using Deep Learning, Singular Value Decomposition, and Self-Organizing Map Approaches*”. Tujuan: Untuk meningkatkan akurasi diagnosis diabetes dengan menggunakan DBN, SVD, dan SOM. Metode: DBN, SVD, dan SOM. Persamaan: pemanfaatan pembelajaran mesin, Perbedaan: Fokus pada diagnosis diabetes awal dengan metode yang lebih kompleks[19].

Berbagai penelitian terkait prediksi diabetes telah dilakukan dengan pendekatan pembelajaran mesin. Namun, sebagian besar penelitian tersebut lebih berfokus pada deteksi dini atau usia onset diabetes, bukan pada prediksi komplikasi. Ringkasan penelitian terdahulu disajikan pada Tabel 1.

Tabel 1 Perbandingan penelitian terdahulu

Judul Penelitian	Tujuan	Metode	Hasil	Persamaan	Perbedaan
<i>Predicting Type 2 diabetes onset age using machine learning: A case study in KSA</i>	Memprediksi usia onset diabetes tipe 2 di Arab Saudi	MLR, ANN, RF, SVR, DTR	MLR & RF memiliki akurasi terbaik, faktor utama: trigliserida, kolesterol, HDL, ferritin, BMI, tekanan darah	Sama-sama menggunakan pembelajaran mesin	Fokus pada usia onset, bukan komplikasi
<i>A diabetes prediction model based on Boruta feature selection and ensemble learning</i>	Meningkatkan akurasi deteksi dini diabetes	Boruta, K-Means++, stacking ensemble	Akurasi 98%	Sama-sama menggunakan seleksi fitur	Fokus pada deteksi dini, metode ensemble kompleks
<i>Advancing diabetes prediction with a progressive self-transfer learning framework for discrete time series data</i>	Mengembangkan prediksi diabetes berbasis data time series	BRITS, LASSO, transfer learning	Peningkatan AUC dan F1 score	Sama-sama memakai algoritma prediktif	Menggunakan deep learning, berbasis time series
<i>A Combined Method for Diabetes Mellitus Diagnosis Using Deep Learning, SVD, and SOM</i>	Meningkatkan akurasi diagnosis diabetes	DBN, SVD, SOM	Akurasi meningkat dengan metode gabungan	Sama-sama memanfaatkan machine learning	Fokus pada diagnosis awal, metode deep learning yang kompleks

Berdasarkan penelitian terdahulu tersebut banyak menggunakan algoritma kompleks seperti **deep learning** dan **ensemble learning**, yang membutuhkan sumber daya komputasi lebih tinggi. Meskipun akurasinya baik, metode tersebut kurang sesuai untuk diterapkan pada klinik pratama yang umumnya memiliki keterbatasan infrastruktur teknologi. Sebaliknya, integrasi Naive Bayes dan Information Gain memiliki beberapa keunggulan yaitu (1) Lebih ringan secara komputasi karena Naive Bayes bekerja dengan kompleksitas linear terhadap jumlah fitur, berbeda dengan deep learning yang membutuhkan komputasi matriks berulang, (2) Lebih hemat memori karena hanya menyimpan probabilitas tiap fitur, bukan bobot jaringan atau model ensemble yang besar, (3) Lebih mudah diimplementasikan karena tidak memerlukan perangkat keras khusus (GPU/TPU), sehingga cukup dijalankan di komputer standar dengan basis data sederhana [20].

Kebaruan penelitian ini terletak pada pengembangan model prediksi komplikasi diabetes yang mengintegrasikan Naive Bayes dan Information Gain secara adaptif. Pendekatan ini belum banyak diterapkan dalam konteks prediksi komplikasi di fasilitas kesehatan primer. Dengan demikian, penelitian ini menawarkan solusi yang lebih praktis, efisien, dan relevan untuk digunakan langsung di klinik pratama.

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain cross-sectional untuk mengembangkan model prediksi komplikasi diabetes melalui Integrasi metode Naive Bayes adaptif dan Information Gain. Adapun Tahapan Diagram Alur Penelitian ini seperti disajikan pada Gambar 1.



Gambar 1 Diagram alur penelitian

1. Pengumpulan dan Preprocessing Data

Penelitian ini mengambil data primer rekam medis pasien diabetes dari Klinik Pratama Surakarta. Data yang dikumpulkan mencakup informasi demografi (usia, jenis kelamin), faktor klinis (HbA1c, tekanan darah, profil lipid, kadar glukosa puasa, BMI), faktor non-klinis (riwayat keluarga, kepatuhan pengobatan, pola makan, aktivitas fisik), dan status komplikasi[21]-[22].

Kemudian dilakukan identifikasi faktor risiko yang signifikan dalam prediksi komplikasi diabetes tahapnya berikut:

- a. Ekstraksi Data Rekam Medis: Data dikumpulkan dari sistem informasi kesehatan Klinik Pratama Surakarta dengan memperhatikan etika dan kerahasiaan.
- b. Analisis Faktor Klinis dan Non-Klinis: Variabel yang dianalisis meliputi faktor klinis dan non-klinis sesuai dengan panduan American Diabetes Association.
- c. Analisis Statistik Deskriptif: Untuk menggambarkan karakteristik populasi pasien dan distribusi faktor risiko.
- d. Analisis Univariat dan Bivariat: Menilai hubungan antara faktor risiko dengan komplikasi diabetes[23].

Proses preprocessing mencakup: *pembersihan data, penanganan missing value, normalisasi, dan transformasi data*. Analisa data dilakukan dengan Tools Software Matlab.

<http://sistemasi.ftik.unisi.ac.id>

2. Implementasi Algoritma Metode Naive Bayes

Pada bagian ini dilakukan Implementasi dari Algoritma Metode Naive Bayes untuk memprediksi deteksi komplikasi diabetes. Dataset yang diambil berdasarkan data dari Klinik Pratama yang telah dibersihkan. Pada implementasi model ini diambil sampel data sebanyak 20 data dengan atribut seperti : Usia, Jenis_Kelamin, Kadar Gula Darah, Tekanan Darah, Riwayat Diabetes, Aktivitas Fisik, Pola Makan. Sedangkan label target komplikasi dengan nilai (Ya dan Tidak). Implementasi dilakukan dengan prosentase data training yaitu 60% dan data testing 40%. Hal ini bertujuan untuk mengukur tingkat kualitas akurasi dataset dalam pengembangan model deteksi dini komplikasi diabetes[25].

3. Seleksi Fitur dengan Information Gain

Pada tahap ini, Metode Information Gain digunakan untuk mengidentifikasi fitur-fitur yang relevan dalam memprediksi komplikasi diabetes. Fitur yang paling relevan dihitung berdasarkan kontribusi informasi setiap fitur terhadap target prediksi[24]. Adapun langkahnya sebagai berikut:

- a. Preprocessing Data: Pembersihan, penanganan missing value, normalisasi, dan transformasi data.
- b. Penghitungan Information Gain: Mengukur kontribusi setiap fitur menggunakan algoritma yang dioptimalkan.
- c. Penentuan Threshold: Menetapkan nilai threshold untuk seleksi fitur.
- d. Reduksi Dimensi: Mengurangi dimensi data dengan mengeliminasi fitur yang kurang relevan.
- e. Validasi Seleksi Fitur: Menggunakan validasi silang untuk menghindari overfitting.

4. Integrasi Metode Naive Bayes dan Information Gain

Pada tahap ini, dikembangkan framework integrasi adaptif yang mengoptimalkan penggunaan kedua metode, yaitu Naive Bayes dan Information Gain[26]. Tahapan untuk integrasi meliputi:

- a. Pengembangan Framework Integrasi: Merancang framework untuk menggabungkan seleksi fitur dan model klasifikasi Naive Bayes[27].
- b. Implementasi Pendekatan Ensemble: Mengintegrasikan hasil beberapa model Naive Bayes dengan fitur yang berbeda[28].
- c. Optimasi Pembobotan Fitur: Mengembangkan mekanisme pembobotan fitur berdasarkan nilai Information Gain untuk meningkatkan akurasi prediksi[29].
- d. Mekanisme Umpan Balik: Mekanisme untuk menyesuaikan parameter berdasarkan hasil validasi.

5. Uji Validasi Model

Model dievaluasi dilakukan pengujian sistem dengan metode confusion matrik untuk mendapatkan tingkat akurasi model. Pengujian ini dilakukan untuk melihat konsistensi dataset yang telah dikembangkan sehingga akhirnya dapat terbentuk suatu rule model yang dapat digunakan dalam melakukan proses prediksi komplikasi diabetes dengan berdasarkan algoritma naive bayes + information gain.

6. Pengembangan Prototipe Sistem Prediksi

Prototipe sistem prediksi komplikasi diabetes dikembangkan dengan antarmuka user-friendly. Prototipe dibuat dengan bahasa PHP dan Database MySQL[30]. Tools Pengembangan sistem yaitu: Dreamweaver, Rapidminer dan Matlab dengan tahap berikut:

- a. **Pengembangan Prototipe:** Merancang prototipe yang mudah digunakan oleh tenaga kesehatan.
- b. **Pengujian Usability:** Menguji prototipe dengan pengguna akhir untuk memastikan kemudahan penggunaan.
- c. **Implementasi Pilot:** Menerapkan sistem di beberapa klinik untuk mengevaluasi efektivitasnya dalam setting klinis.
- d. **Dokumentasi dan Pelatihan:** Menyediakan dokumentasi dan pelatihan untuk tenaga kesehatan.

7. Evaluasi Komparasi Algoritma

Evaluasi komparasi ini bertujuan untuk melakukan perbandingan hasil akurasi antara algoritma naive bayes dengan algoritma naive bayes + information.

4 Hasil dan Pembahasan

Pada pembahasan penelitian ini, mengenai pengembangan model prediksi komplikasi diabetes menggunakan integrasi adaptif metode Naive Bayes dan Information Gain yang akan dibahas secara rinci. Penelitian ini melibatkan langkah-langkah pengumpulan data, preprocessing, seleksi fitur, pengembangan model, serta evaluasi dan validasi model.

A. Pengumpulan dan Preprocessing Data

Data yang digunakan dalam penelitian ini diambil dari rekam medis pasien diabetes di Klinik Pratama Surakarta, yang mencakup informasi demografi, faktor klinis, dan faktor non-klinis. Data ini diproses melalui beberapa tahapan seperti pembersihan data, penanganan missing values, dan normalisasi. Berdasarkan yang telah dikumpulkan, peneliti mengambil beberapa sampel data sebanyak 20 dataset seperti disajikan pada Tabel 1.

Tabel 1 Dataset diabetes sebelum normalisasi

ID	Usia	Jenis Kelamin	Kadar Gula Darah (mg/dL)	HbA1c (%)	Tekanan Darah (mmHg)	Riwayat Keluarga Diabetes	Aktivitas Fisik (minggu)	Pola Makan	Komplikasi
1	45	Laki-laki	150	7.5	140/90	Ya	5	Tinggi Gula	1
2	60	Perempuan	120	6.8	130/85	Tidak	2	Seimbang	0
3	38	Laki-laki	170	8.1	145/95	Ya	1	Gula Tinggi	1
4	50	Perempuan	110	6.5	125/80	Ya	4	Seimbang	0
5	55	Laki-laki	180	9.0	150/100	Ya	0	Tinggi Lemak	1
6	42	Perempuan	140	7.2	135/88	Tidak	3	Seimbang	0
7	67	Laki-laki	160	7.8	160/105	Ya	1	Tinggi Gula	1
8	30	Perempuan	125	6.3	120/78	Tidak	5	Seimbang	0
9	48	Laki-laki	200	9.5	132/85	Ya	0	Gula Tinggi	1
10	37	Perempuan	110	6.1	130/85	Tidak	4	Seimbang	0
11	66	Laki-laki	185	8.0	145/92	Ya	2	Tinggi Gula	1
12	41	Perempuan	130	6.6	130/86	Tidak	4	Seimbang	0
13	53	Laki-laki	131	7.9	140/94	Ya	3	Gula Tinggi	1
14	61	Perempuan	115	6.4	120/80	Ya	5	Seimbang	0
15	56	Laki-laki	190	8.5	150/98	Ya	5	Tinggi Lemak	1
16	29	Perempuan	140	7.2	130/84	Tidak	4	Seimbang	0
17	47	Laki-laki	165	7.7	145/93	Ya	2	Tinggi Gula	1
18	52	Perempuan	120	6.3	135/90	Tidak	3	Seimbang	0
19	59	Laki-laki	180	8.2	150/95	Ya	0	Gula Tinggi	1
20	33	Perempuan	110	6.0	125/82	Tidak	5	Seimbang	0

Berdasarkan dataset dari Tabel 1 tersebut masih terdapat nilai number dan belum dikategorikan, kemudian pada tahapan ini dilakukan proses pengkategorian data pada atribut Usia, Kadar Gula Darah (mg/dL), HbA1c (%), Tekanan Darah (mmHg) dan Aktivitas Fisik (minggu). Pada bagian label/target dikegerikan langsung yakni 1 = Ya dan 0 = Tidak. Berikut konversi kategori atribut:

Konversi Kategori => Usia

- ✓ Anak Muda: Usia 18 - 35 tahun
- ✓ Dewasa Tengah: Usia 36 - 50 tahun
- ✓ Lansia Awal: Usia 51 - 65 tahun
- ✓ Lansia: Usia ≥ 66 tahun

Konversi Kategori => Kadar Gula Darah (mg/dL)

- ✓ Rendah: ≤ 125
- ✓ Sedang: 126–150
- ✓ Tinggi: 151–175
- ✓ Sangat Tinggi: ≥ 176

Konversi Kategori => HbA1c (%)

- ✓ Normal: $< 5.7\%$
- ✓ Pradiabetes: $5.7\% - 6.4\%$
- ✓ Diabetes: $\geq 6.5\%$

Konversi Kategori => Tekanan Darah (mmHg)

- ✓ Normal: Sistolik < 120 dan Diastolik < 80
- ✓ Prehipertensi: Sistolik 120-139 atau Diastolik 80-89
- ✓ Hipertensi: Sistolik $\geq$ 140 atau Diastolik $\geq$ 90

Catatan : Pada Atribut HbA1c (%) & Tekanan Darah (mmHg) memiliki kategori yang sama maka diambil salah satu saja.

Konversi Kategori => Aktivitas Fisik (minggu)

- ✓ Tidak Aktif: 0 aktivitas fisik → Tidak ada aktivitas fisik yang dilakukan
- ✓ Sedikit Aktif: 1-2 aktivitas fisik per minggu → Aktivitas fisik minimal
- ✓ Cukup Aktif: 3-4 aktivitas fisik per minggu → Aktivitas fisik yang cukup
- ✓ Sangat Aktif: 5 aktivitas fisik atau lebih per minggu → Aktivitas fisik tinggi, sehat

Berdasarkan konversi kategori tersebut maka dataset awal dari Tabel 1 dapat di normalisasi dengan penyesuaian seperti pada Tabel 2.

Tabel 2 Dataset diabetes

ID	Usia	Jenis Kelamin	Kadar Gula Darah	Tekanan Darah	Riwayat Diabetes	Aktivitas Fisik	Pola Makan	Komplikasi
1	Dewasa Tengah	Laki-laki	Sedang	Diabetes	Ya	Sangat Aktif	Tinggi Gula	Ya
2	Lansia Awal	Perempuan	Rendah	Pradiabetes	Tidak	Sedikit Aktif	Seimbang	Tidak
3	Dewasa Tengah	Laki-laki	Tinggi	Diabetes	Ya	Tidak Aktif	Gula Tinggi	Ya
4	Dewasa Tengah	Perempuan	Rendah	Pradiabetes	Ya	Cukup Aktif	Seimbang	Tidak
5	Lansia Awal	Laki-laki	Sangat Tinggi	Diabetes	Ya	Tidak Aktif	Tinggi Lemak	Tidak
6	Dewasa Tengah	Perempuan	Sedang	Pradiabetes	Tidak	Cukup Aktif	Seimbang	Ya
7	Lansia	Laki-laki	Tinggi	Diabetes	Ya	Tidak Aktif	Tinggi Gula	Ya
8	Anak Muda	Perempuan	Rendah	Pradiabetes	Tidak	Sangat Aktif	Seimbang	Tidak
9	Dewasa Tengah	Laki-laki	Sangat Tinggi	Pradiabetes	Ya	Tidak Aktif	Gula Tinggi	Ya
10	Dewasa Tengah	Perempuan	Rendah	Pradiabetes	Tidak	Cukup Aktif	Seimbang	Tidak
11	Lansia Awal	Laki-laki	Sangat Tinggi	Diabetes	Ya	Sedikit Aktif	Tinggi Gula	Ya
12	Dewasa Tengah	Perempuan	Sedang	Pradiabetes	Tidak	Cukup Aktif	Seimbang	Tidak
13	Dewasa Tengah	Laki-laki	Sedang	Diabetes	Ya	Cukup Aktif	Gula Tinggi	Tidak
14	Lansia Awal	Perempuan	Rendah	Pradiabetes	Ya	Sangat Aktif	Seimbang	Tidak
15	Lansia Awal	Laki-laki	Sangat Tinggi	Diabetes	Ya	Sangat Aktif	Tinggi Lemak	Ya
16	Anak Muda	Perempuan	Sedang	Pradiabetes	Tidak	Cukup Aktif	Seimbang	Tidak
17	Dewasa Tengah	Laki-laki	Tinggi	Diabetes	Ya	Sedikit Aktif	Tinggi Gula	Ya
18	Dewasa Tengah	Perempuan	Rendah	Pradiabetes	Tidak	Cukup Aktif	Seimbang	Tidak
19	Lansia Awal	Laki-laki	Sangat Tinggi	Diabetes	Ya	Tidak Aktif	Gula Tinggi	Ya
20	Anak Muda	Perempuan	Rendah	Pradiabetes	Tidak	Sangat Aktif	Seimbang	Tidak

Pada Tabel 2 merupakan dataset yang telah dinormalisasi sebagai sampel dalam pengujian algoritma. Hal ini bertujuan untuk membuat suatu desain model Prediksi Komplikasi Diabetes.

B. Implementasi Algoritma Naïve Bayes

Pada tahapan ini dilakukan proses testing pengujian sistem untuk memprediksi apakah seseorang akan mengalami komplikasi (Komplikasi = Ya/Tidak) berdasarkan 7 atribut: Usia, Jenis Kelamin, Kadar Gula Darah, Tekanan Darah, Riwayat Diabetes, Aktivitas Fisik dan Pola Makan.

Langkah Perhitungan (Manual Naïve Bayes)

Misalkan disini peneliti ingin memprediksi data pasien baru dengan mengambil dataset dari ID : 13 diperoleh data testing seperti pada Tabel 3.

Tabel 3 Contoh data testing

Atribut	Nilai
Usia	Dewasa Tengah
Jenis_Kelamin	Laki-laki
Kadar_Gula_Darah	Sedang
Tekanan_Darah	Diabetes
Riwayat_Diabetes	Ya
Aktivitas_Fisik	Sangat Aktif
Pola_Makan	Tinggi Gula

Berdasarkan dataset diperoleh hasil Komplikasi = Tidak, namun peneliti akan melakukan eksperimen dengan perhitungan Algoritma Naïve Bayes berdasarkan 7 variabel dengan tahapan berikut:

1. Hitung Probabilitas Kelas

Jumlah data total: 20

- Jumlah Komplikasi = Ya: 9
- Jumlah Komplikasi = Tidak: 11

$P(Ya)=9/20=0,45$ dan $P(Tidak)=11/20 = 0,55$

2. Hitung Probabilitas Tiap Atribut Berdasarkan Kelas

Contoh Kasus ID : 13 Untuk menghitung $P(\text{Usia} = \text{Dewasa Tengah} \mid \text{Komplikasi} = \text{Ya}) \rightarrow$ Dari 9 data "Komplikasi = Ya", berapa yang Usia = Dewasa Tengah?

Lihat data: ada 5 yang Dewasa Tengah dari total 9 $\rightarrow P(\text{DewasaTengah} \mid Ya) = 5/9 = 0,55$ Lakukan hal serupa untuk semua atribut:

a. $P(x \mid \text{Komplikasi} = \text{Ya})$

Tabel 4 Rekapitulasi perhitungan atribut (ya)

Atribut	Nilai	$P(x \mid Ya)$
Usia	Dewasa Tengah	$5/9 = 0,55$
Jenis Kelamin	Laki-laki	$5/9 = 0,55$
Kadar Gula Darah	Sedang	$2/9 = 0,22$
Tekanan Darah	Diabetes	$7/9 = 0,77$
Riwayat Diabetes	Ya	$9/9 = 1$
Aktivitas Fisik	Sangat Aktif	$2/9 = 0,22$
Pola Makan	Tinggi Gula	$4/9 = 0,44$

$P(X \mid Ya) = (5/9 \times 5/9 \times 2/9 \times 7/9 \times 1 \times 2/9 \times 4/9) = 0.000431$

$P(Ya \mid X) = P(X \mid Ya) \times P(Ya) = 0.000431 \times 9/20 = 0.000194$

b. $P(x \mid \text{Komplikasi} = \text{Tidak})$

Tabel 5 Rekapitulasi perhitungan atribut (tidak)

Atribut	Nilai	$P(x \mid \text{Tidak})$
Usia	Dewasa Tengah	$6/11 = 0,54$
Jenis Kelamin	Laki-laki	$4/11 = 0,36$
Kadar Gula Darah	Sedang	$3/11 = 0,27$
Tekanan Darah	Diabetes	$2/11 = 0,18$
Riwayat Diabetes	Ya	$2/11 = 0,18$
Aktivitas Fisik	Sangat Aktif	$3/11 = 0,27$
Pola Makan	Tinggi Gula	$1/11 = 0,09$

$P(X \mid \text{Tidak}) = (6/11 \times 4/11 \times 3/11 \times 2/11 \times 2/11 \times 3/11 \times 1/11) = 0.0000123$

$P(\text{Tidak} \mid X) = P(X \mid \text{Tidak}) \times P(\text{Tidak}) = 0.0000123 \times 11/20 = 0.00000676$

Hasil Akhir

Karena $\Rightarrow P(Ya \mid X) > P(\text{Tidak} \mid X)$, Maka **Prediksi: Komplikasi = Ya**

Kesimpulan : Berdasarkan hasil implementasi algoritma naïve bayes dari dataset diatas dengan 7 variabel diperoleh prediksi komplikasi = Ya (Error) tidak sesuai dataset. Untuk itu pada pengembangan algoritma ini dilakukan penambahan implementasi algoritma information gain yang berguna untuk menentukan variabel terbaik yang mampu memberikan hasil prediksi dengan tepat.

C. Seleksi Fitur dengan Information Gain

Information Gain adalah sebuah konsep dalam machine learning dan khususnya dalam algoritma pohon keputusan. Secara sederhana, Information Gain mengukur seberapa banyak informasi yang diperoleh dari sebuah fitur dalam konteks klasifikasi data. Semakin tinggi Information Gain suatu fitur, semakin baik fitur tersebut dalam memisahkan atau mengklasifikasikan data ke dalam kelas-kelas yang berbeda. Tujuan dari Algoritma Algoritma Information Gain adalah untuk menentukan variabel terbaik atau optimal.

Langkah Perhitungan Algoritma Information Gain sebagai berikut:

1. Hitung Entropy Awal (Entropy(S))

Target Komplikasi:

- ✓ Jumlah total data: 20
- ✓ Ya: 9
- ✓ Tidak: 11

$$Entropy(S) = - \left(\frac{9}{20} \cdot \log_2 \frac{9}{20} + \frac{11}{20} \cdot \log_2 \frac{11}{20} \right) \approx - (0.45 \cdot (-1.152) + 0.55 \cdot (-0.858)) \approx 0.991$$

2. Hitung Information Gain untuk Setiap Atribut

1). Atribut Usia

Nilai unik: Dewasa Tengah (9), Lansia Awal (6), Lansia (1), Anak Muda (4) Hitung entropy masing-masing:

- ✓ **Dewasa Tengah (9 data)** → Ya: 4, Tidak: 5

$$Entropy = - \left(\frac{4}{9} \log_2 \frac{4}{9} + \frac{5}{9} \log_2 \frac{5}{9} \right) \approx 0.991$$

- ✓ **Lansia Awal (6 data)** → Ya: 2, Tidak: 4

$$Entropy = - \left(\frac{2}{6} \log_2 \frac{2}{6} + \frac{4}{6} \log_2 \frac{4}{6} \right) \approx 0.918$$

- ✓ **Lansia (1 data)** → Ya: 1 → Entropy = 0

- ✓ **Anak Muda (4 data)** → Ya: 0, Tidak: 4 → Entropy = 0

$$IG(Usia) = Entropy(S) - \sum \left(\frac{|Sv|}{|S|} \cdot Entropy(Sv) \right)$$

$$IG(Usia) = 0.991 - \left(\frac{9}{20} \cdot 0.991 + \frac{6}{20} \cdot 0.918 + \frac{1}{20} \cdot 0 + \frac{4}{20} \cdot 0 \right)$$

$$IG(Usia) = 0.991 - (0.446 + 0.275 + 0 + 0) = 0.991 - 0.721 = 0.270$$

2). Atribut Jenis_Kelamin

Nilai: Laki-laki (10), Perempuan (10)

- ✓ **Laki-laki**: Ya: 6, Tidak: 4 → Entropy ≈ 0.971
- ✓ **Perempuan**: Ya: 3, Tidak: 7 → Entropy ≈ 0.881

$$IG(Jenis_{Kelamin}) = 0.991 - \left(\frac{10}{20} \cdot 0.971 + \frac{10}{20} \cdot 0.881 \right) = 0.991 - (0.4855 + 0.4405) = 0.065$$

3). Atribut Kadar_Gula_Darah

Pada Nilai Atribut Kadar Gula Darah : Sedang, Rendah, Tinggi, Sangat Tinggi, hasilnya seperti disajikan pada Tabel 6.

Tabel 6 Rekapitulasi hasil entropy kadar gula darah

Nilai	Total	Ya	Tidak	Entropy
Sedang	4	2	2	1.000
Rendah	7	1	6	0.592
Tinggi	3	3	0	0.000
Sangat Tinggi	6	3	3	1.000

$$IG = 0.991 - \left(\frac{4}{20} \cdot 1 + \frac{7}{20} \cdot 0.592 + \frac{3}{20} \cdot 0 + \frac{6}{20} \cdot 1 \right) = 0.991 - (0.2 + 0.207 + 0 + 0.3) = 0.991 - 0.707 = 0.284$$

4). Atribut Tekanan_Darah

Pada Nilai Atribut Tekanan Darah : Diabetes (10), Prediabetes (10), seperti pada Tabel 7.

Tabel 7 Rekapitulasi hasil entropy tekanan darah

Nilai	Ya	Tidak	Entropy
Diabetes	7	3	0.881
Prediabetes	2	8	0.722

$$IG = 0.991 - \left(\frac{10}{20} \cdot 0.881 + \frac{10}{20} \cdot 0.722 \right) = 0.991 - (0.4405 + 0.361) = 0.189$$

5). Atribut Riwayat_Diabetes

Pada Nilai Atribut Riwayat Diabete yaitu : Ya (11), Tidak (9), hasilnya seperti pada Tabel 8.

Tabel 8 Rekapitulasi hasil entropy riwayat diabetes

Nilai	Ya	Tidak	Entropy
Ya	7	4	0.985
Tidak	2	7	0.764

$$IG=0.991-(0.55 \times 0.985 + 0.45 \times 0.764) = 0.991 - (0.5418 + 0.3438) = 0.105$$

6). Atribut Aktivitas_Fisik

Pada Atribut Aktivitas Fisik, terlalu banyak nilai unik → kita kelompokkan nilai-nilainya:
Kelompokkan sebagai:

- ✓ Aktif (Cukup Aktif, Sangat Aktif)
- ✓ Tidak Aktif (Tidak Aktif, Sedikit Aktif)

Tabel 9 Rekapitulasi hasil entropy riwayat aktivitas fisik

Kategori	Total	Ya	Tidak	Entropy
Aktif	10	2	8	0.722
Tidak Aktif	10	7	3	0.881

$$IG=0.991-(0.5 \times 0.722 + 0.5 \times 0.881) = 0.991 - 0.8015 = 0.189$$

7). Atribut Pola_Makan

Pada Atribut Pola maka dapat di kelompokkan sebagai:

- ✓ **Tinggi Gula / Lemak**
- ✓ **Seimbang**

Tabel 10 Rekapitulasi hasil entropy pola makan

Kategori	Total	Ya	Tidak	Entropy
Tinggi	9	6	3	0.918
Seimbang	11	3	8	0.845

$$IG=0.991-(0.45 \times 0.918 + 0.55 \times 0.845) = 0.991 - (0.413 + 0.464) = 0.114$$

Rangkuman Information Gain

Berdasarkan perhitungan Algoritma Information Gain maka dapat diambil kesimpulan rekapitulasi seperti disajikan pada Tabel 11.

Tabel 11 Rekapitulasi hasil rank variabel dengan IG

Atribut	Information Gain	Rank
Kadar_Gula_Darah	0.284	1
Usia	0.270	2
Tekanan Darah	0.189	3
Aktivitas Fisik	0.189	4
Pola Makan	0.114	5
Riwayat Diabetes	0.105	6
Jenis_Kelamin	0.065	7

Setelah diperoleh hasil variabel optimal dari Tabel 11. Maka peneliti melakukan eksperimen dengan mengambil 4 variabel terbaik yaitu Kadar Gula Darah, Usia, Tekanan Darah dan Aktivitas Fisik.

D. Integrasi Metode Naive Bayes dan Information Gain

Pada implementasi studi kasus ini menggunakan Algoritma Naïve Bayes + Information Gain yaitu dengan menggunakan 4 variabel optimal seperti: Kadar Gula Darah, Usia, Tekanan Darah

dan Aktivitas Fisik. Pada pengujian model ini dilakukan dengan data uji kasus yang sama pada ID : 13 sebagai berikut:

Data Kasus Uji Testing sebagai berikut:

Tabel 12 Data testing uji baru

Atribut	Nilai
Usia	=> Dewasa Tengah
Kadar_Gula_Darah	=> Sedang
Tekanan_Darah	=> Diabetes
Aktivitas_Fisik	=> Sangat Aktif

Langkah 1: Hitung Probabilitas Dasar

Total data = 20

✓ Komplikasi = Ya: 9

✓ Komplikasi = Tidak: 11

$$P(Ya) = \frac{9}{20}, \quad P(Tidak) = \frac{11}{20}$$

Langkah 2: Hitung Probabilitas Tiap Atribut Berdasarkan Kelas

Komplikasi = Ya

Tabel 13 Hasil nilai probabilitas atribut

Atribut	Nilai	Jumlah Ya (9)	P(x Ya)
Usia	Dewasa Tengah	5	5/9
Kadar Gula Darah	Sedang	2	2/9
Tekanan Darah	Diabetes	7	7/9
Aktivitas Fisik	Sangat Aktif	2	2/9

$$P(X|Ya) = \frac{5}{9} \cdot \frac{2}{9} \cdot \frac{7}{9} \cdot \frac{2}{9} \approx 0.000431$$

$$P(Ya|X) = P(X|Ya) \cdot P(Ya) = 0.000431 \cdot \frac{9}{20} \approx 0.000194$$

Komplikasi = Tidak

Tabel 14 Hasil rekapitulasi nilai atribut pada komplikasi “tidak”

Atribut	Nilai	Jumlah Tidak (11)	P(x Tidak)
Usia	Dewasa Tengah	4	4/11
Kadar Gula Darah	Sedang	2	2/11
Tekanan Darah	Diabetes	3	3/11
Aktivitas Fisik	Sangat Aktif	3	3/11

$$P(X|Tidak) = \frac{4}{11} \cdot \frac{2}{11} \cdot \frac{3}{11} \cdot \frac{3}{11} \approx 0.000541$$

$$P(Tidak|X) = P(X|Tidak) \cdot P(Tidak) = 0.000541 \cdot \frac{11}{20} \approx 0.000297$$

HASIL AKHIR

$$P(Ya|X) = 0.000194, \quad P(Tidak|X) = 0.000297$$

Karena:

$$P(Tidak|X) > P(Ya|X)$$

Hasil Prediksi: Komplikasi = Tidak

Berdasarkan hasil perhitungan manual dengan algoritma naïve bayes + information gain maka diperoleh hasil komplikasi = “Tidak” hasil ini dinyatakan valid sesuai dengan dataset ID:13, dengan demikian pengembangan Algoritma naïve bayes + information gain telah berjalan dengan baik.

E. Uji Validasi Model

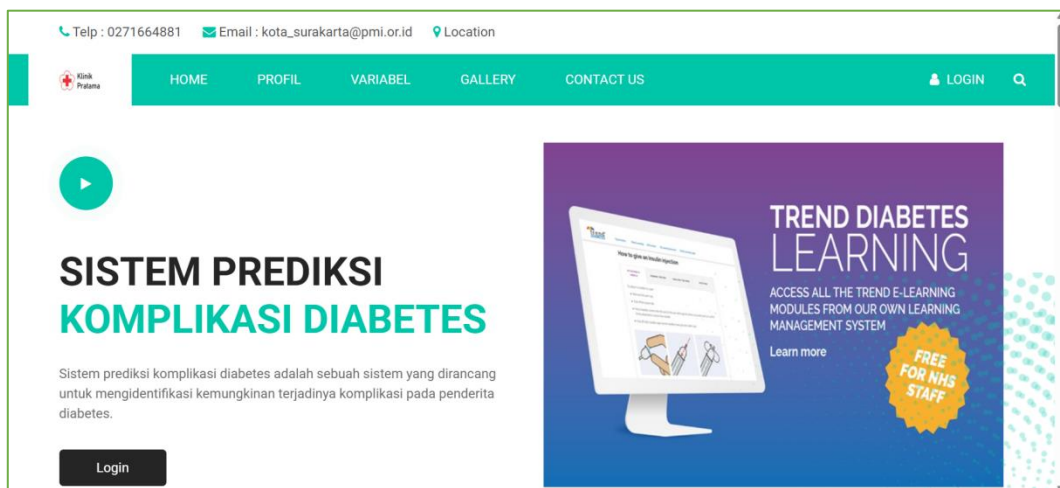
Pengujian Model dilakukan dengan dengan metode confusion matrik untuk mendapatkan tingkat akurasi model. Pengujian ini dilakukan pada prediksi dengan algoritma naïve bayes dan integrasi algortima naïve bayes + informatioan gain. Adapun rekapitulasi hasil pengujian model seperti disajikan pada Tabel 15.

Tabel 15 Komparasi uji validasi model algoritma

Algoritma Naïve Bayes			Algoritma Naïve Bayes + Information Gain		
	Ya	Tidak		Ya	Tidak
Ya	2	1	Ya	3	0
Tida	1	4	Tidak	1	4
Hasil Akurasi: $TP+TN/(TP+TN+FP+FN)$: 75%			Hasil Akurasi: $TP+TN/(TP+TN+FP+FN)$: 87.5%		

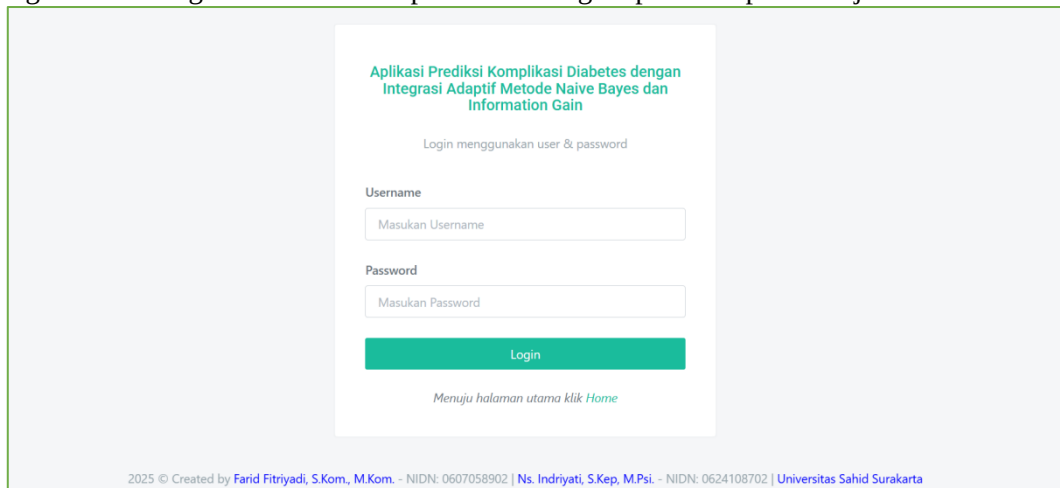
F. Pengembangan Prototipe Sistem Prediksi

Pada pengembangan sistem ini dibuat dengan menggunakan bahasa pemrograman php dan database mysql sehingga terbentuknya desain aplikasi sebagai prototipe dari pengembangan sistem seperti disajikan pada Gambar 2.



Gambar 2 Tampilan halaman awal aplikasi

Pada Aplikasi ini dilengkapi dengan beberapa menu utama: Frond End => Home | Profil | Variabel | Gallery | Contact Us. Kemudian pada bagian Back End terdapat dashboard untuk proses prediksi yang dilakukan dengan import file dari dataset yang telah dirapikan sesuai dengan atribut diperlukan. Adapun tampilan halaman back end aplikasi dapat dilakukan dengan mengklik menu login maka akan tampil halaman login aplikasi seperti disajikan di Gambar 2.



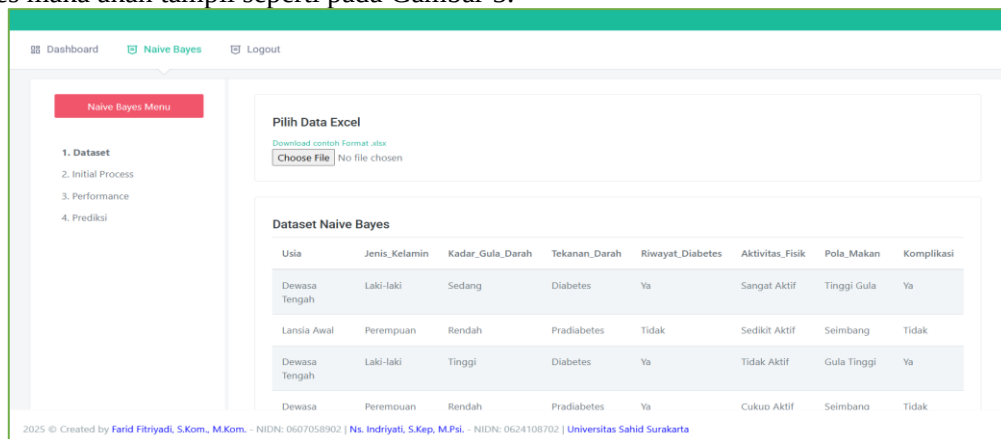
Gambar 3 Tampilan halaman login aplikasi

Pada Implementasi Aplikasi ini, untuk login dapat dilakukan dengan memasukkan username : admin dan password: admin. Bila telah berhasil login maka akan tampil halaman dashboard admin seperti disajikan pada Gambar 4.



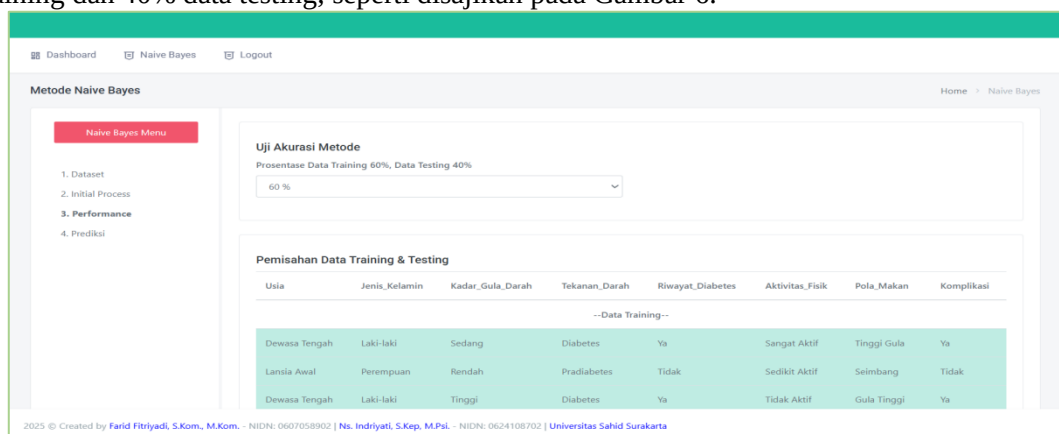
Gambar 4 Tampilan halaman dashbord administrator

Untuk melakukan proses prediksi maka dapat dilakukan dengan mengklik tombol menu Naïve Bayes maka akan tampil seperti pada Gambar 5.



Gambar 5 Tampilan halaman import dataset

Pada tahap selanjutnya Menu Initalisasi Proses => Untuk membagi antara dataset atribut dengan label/kelas atau target, kemudian pada menu Performance digunakan untuk membagi presentase antara data training dengan data testing. Pada uji eksperimen ini dibagi 60% data training dan 40% data testing, seperti disajikan pada Gambar 6.



Gambar 6 Uji akurasi metode

Berdasarkan dataset dengan 7 variabel maka diperoleh tingkat akurasi seperti disajikan pada Gambar 7.

Dashboard

Naive Bayes

Logout

Anak Muda	Perempuan	Sedang	Pradiabetes	Tidak	Cukup Aktif	Seimbang	Tidak	Tidak
Dewasa Tengah	Laki-laki	Tinggi	Diabetes	Ya	Sedikit Aktif	Tinggi Gula	Ya	Ya
Dewasa Tengah	Perempuan	Rendah	Pradiabetes	Tidak	Cukup Aktif	Seimbang	Tidak	Tidak
Lansia Awal	Laki-laki	Sangat Tinggi	Diabetes	Ya	Tidak Aktif	Gula Tinggi	Ya	Ya
Anak Muda	Perempuan	Rendah	Pradiabetes	Tidak	Sangat Aktif	Seimbang	Tidak	Tidak

	Ya	Tidak
Ya	2	1
Tidak	1	4

Hasil Akurasi : $TP+TN/(TP+TN+FP+FN)$: 75%

2025 © Created by Farid Fitriyadi, S.Kom., M.Kom. - NIDN: 0607058902 | Ns. Indriyati, S.Kep. M.Psi. - NIDN: 0624108702 | Universitas Sahid Surakarta

Gambar 7 Hasil akurasi model dengan 7 variabel

Sedangkan uji dataset dengan algoritma naïve bayes + information gain dengan 4 variabel diperoleh akurasi tinggi seperti disajikan pada Gambar 8.

Dashboard

Naive Bayes

Logout

Lansia Awal	Sangat Tinggi	Diabetes	Sangat Aktif	Ya	Ya
Anak Muda	Sedang	Pradiabetes	Cukup Aktif	Tidak	Tidak
Dewasa Tengah	Tinggi	Diabetes	Sedikit Aktif	Ya	Ya
Dewasa Tengah	Rendah	Pradiabetes	Cukup Aktif	Tidak	Tidak
Lansia Awal	Sangat Tinggi	Diabetes	Tidak Aktif	Ya	Ya
Anak Muda	Rendah	Pradiabetes	Sangat Aktif	Tidak	Tidak

	Ya	Tidak
Ya	3	0
Tidak	1	4

Hasil Akurasi : $TP+TN/(TP+TN+FP+FN)$: 87.5%

2025 © Created by Farid Fitriyadi, S.Kom., M.Kom. - NIDN: 0607058902 | Ns. Indriyati, S.Kep. M.Psi. - NIDN: 0624108702 | Universitas Sahid Surakarta

Gambar 8 Hasil akurasi model dengan 4 variabel

Pada tahap testing pengujian sistem dari model yang telah dibuat, maka peneliti menggunakan data testing dari ID:13 sesuai dengan pembahasan diatas diperoleh hasil seperti pada gambar 8.

Dashboard	Naive Bayes	Logout
1. Dataset 2. Initial Process 3. Performance 4. Prediksi	Nama Ahmad Usia Dewasa Tengah Jenis_Kelamin Laki-laki Kadar_Gula_Darah Sedang Tekanan_Darah Diabetes Risikot_Diabetes Ya Aktivitas_Fisik Sangat Aktif Pola_Makan Tinggi Gula Hasil Prediksi Ya	Total Label Ya : 9 (0.0033719641503008) Tidak : 11 (0) Total : 20 Hasil Prediksi Array ([Tidak] => 0 [Ya] => 0.0033719641503008) Tidak : 0 Ya : 0.0033719641503008 Usia : Dewasa Tengah Ya : 3, Tidak : 5 Dewasa Tengah Ya : 5 S/N: 0.5555555555555556 Tidak : 5 S/N: 0.4545454545454545 Lansia Awal Ya : 3 S/N: 0.3333333333333333 Tidak : 3 S/N: 0.2727272727272727

2025 © Created by Farid Fitriyadi, S.Kom., M.Kom. - NIDN: 0607058902 | Ns. Indriyati, S.Kep. M.Psi. - NIDN: 0624108702 | Universitas Sahid Surakarta

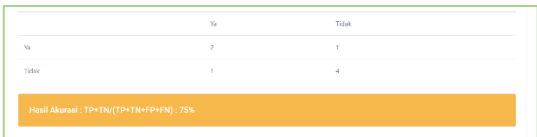
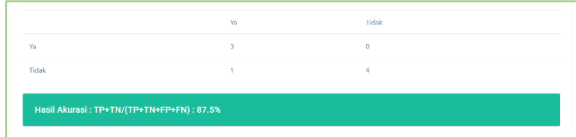
Gambar 9 Implementasi prediksi

Hasil Prediksi : “Ya” ini dinyatakan valid karena sesuai dengan perhitungan manual dari Algoritma Naïve Bayes.

G. Evaluasi Komparasi Algoritma

Evaluasi Akhir dari Optimalisasi Pengembangan Model Prediksi ini yakni melakukan komparasi antara algoritma naïve bayes dengan algoritma naïve bayes + information gain. Hasil komparasi ini bertujuan untuk melihat tingkat akurasi terbaik di antara algoritma tersebut. Pada skenario uji komparasi untuk optimalisasi model dapat dilihat pada Tabel 16.

Tabel 16 Perbandingan pengujian model

Model Algoritma Naïve Bayes	Model Algoritma Naïve Bayes + Information Gain
Jumlah Variabel : 7 Atribut (Usia, Jenis Kelamin, Kadar Gula Darah, Tekanan Darah, Riwayat Diabetes, Aktivitas Fisik, Pola Makan) Akurasi Model diperoleh : 75%	Jumlah Variabel Setelah di seleksi dengan IG : 4 Atribut (Kadar Gula Darah, Usia, Tekanan Darah, Aktivitas Fisik) Akurasi Model diperoleh : 87,5%
	

Hasil Perhitungan Manual vs Program Valid

Hasil Perhitungan Manual vs Program Valid

Berdasarkan hasil perbandingan pengujian model menunjukkan bahwa dengan pengembangan model menggunakan dua algoritma yaitu Naïve Bayes + Information Gain diperoleh Optimalisasi Peningkatan Akurasi dari 75% menjadi 87,5% yaitu meningkat akurasinya 12,5%.

5 Kesimpulan

Penelitian ini menunjukkan bahwa integrasi adaptif metode Naive Bayes dan Information Gain mampu meningkatkan performa prediksi komplikasi diabetes dibandingkan Naive Bayes murni. Hasil uji coba memperlihatkan peningkatan akurasi dari 75% menjadi 87,5% sehingga terdapat peningkatan akurasi optimalisasi model 12,5%. Meskipun demikian, capaian akurasi 87,5% perlu ditafsirkan secara hati-hati karena ukuran sampel yang terbatas (20 data) berpotensi menimbulkan bias. Oleh karena itu, penelitian lanjutan dengan jumlah data yang lebih besar dan beragam sangat diperlukan untuk memastikan konsistensi performa model. Dari sisi implementasi, model ini relatif ringan secara komputasi, hemat memori, dan mudah diintegrasikan dengan sistem informasi berbasis web sederhana. Hal ini menjadikannya sesuai untuk diterapkan di Klinik Pratama yang memiliki keterbatasan sumber daya teknologi. Dengan adanya sistem prediksi ini, tenaga kesehatan dapat terbantu dalam mengidentifikasi pasien berisiko tinggi secara lebih objektif, sehingga intervensi pencegahan dapat dilakukan lebih cepat, terarah, dan efisien. Secara praktis, penelitian ini berkontribusi pada penguatan layanan kesehatan primer melalui penyediaan alat bantu pengambilan keputusan yang sederhana namun efektif. Implikasi jangka panjangnya adalah potensi menurunkan angka komplikasi, mengurangi beban perawatan di rumah sakit, serta mendukung program nasional pengendalian penyakit tidak menular.

Ucapan Terima Kasih

Penulis menyampaikan apresiasi dan terima kasih kepada Direktorat Penelitian dan Pengabdian Masyarakat (DPPM) Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi yang telah memberikan dukungan pendanaan melalui skema Penelitian Dosen Pemula (PDP) Tahun 2025. Ucapan terima kasih juga ditujukan kepada pihak Universitas Sahid Surakarta serta Klinik Pratama Surakarta atas dukungan fasilitas, data, dan kerja sama yang memungkinkan penelitian ini dapat terlaksana dengan baik. Semoga hasil penelitian ini memberikan manfaat bagi pengembangan ilmu pengetahuan dan peningkatan kualitas layanan kesehatan masyarakat.

Referensi

- [1] Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. *IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045*. *Diabetes Res Clin Pract*. 2022;183:1–23. Available from: <https://doi.org/10.1016/j.diabres.2021.109119>
- [2] Aisyah R, Mahmudah N, Candrasari A, Sintowati R, Bestari RS, Rosyidah DU, et al. *Improving Health Profiles and Type 2 Diabetes Mellitus Knowledge of Community Groups in Karanglo Village , Grogol , Sukoharjo , Central Java , Indonesia Through the SEE Method*. 2024;10(4):196–203. Available from: <https://doi.org/10.22146/jpkm.95765>
- [3] Mezil, Sabreen A BA abed. *Complication of Diabetes Mellitus*. *Ann Rom Soc Cell Biol*. 2021;25(3)(3):1546–56. Available from: <https://www.proquest.com/openview/1edb7aa55be175af429f8b2b45ab2bae/1?cbl=2031963&pq-origsite=gscholar>
- [4] Hadi C, Isa M, Sutrisna E, Pangestu M. *Patient Satisfaction and Compliance in One-Day Care Chemotherapy : A Study at Dr . Moewardi Hospital*. 2025;14(April):20–31. Available from: <https://doi.org/10.18196/jmmr.v14i1.531>
- [5] Setyawati R, Astuti A, Utami TP, Adiwijaya S, Hasyim DM. *The Importance of Early Detection in Disease Management the Importance of Early Detection in Disease Management Sekolah Tinggi Ilmu Kesehatan Kesosi , Indonesia Poltekkes Kemenkes Kupang , Indonesia Universitas Esa Unggul , Indonesia Universitas Palangka*. *J World Futur Med Heal Nurs*. 2024;2(1):50–63. Available from: <https://doi.org/10.55849/health.v2i1.692>
- [6] Feng B, Shi J, Huang L, Yang Z, Feng ST, Li J, et al. *Robustly Federated Learning Model for Identifying High-Risk Patients with Postoperative Gastric Cancer Recurrence*. *Nat Commun*. 2024;15(1):1–11. Available from: <https://doi.org/10.1038/s41467-024-44946-4>
- [7] Natarajan K, Baskaran D, Kamalanathan S. *An Adaptive Ensemble Feature Selection Technique for Model-Agnostic Diabetes Prediction*. *Sci Rep*. 2025;15(1):1–12. Available from: <https://doi.org/10.1038/s41598-025-91282-8>
- [8] Khan FA, Zeb K, Al-Rakhami M, Derhab A, Bukhari SAC. *Detection and Prediction of Diabetes using Data Mining: A Comprehensive Review*. *IEEE Access*. 2021;9:43711–35. Available from: <http://doi.org/10.1109/ACCESS.2021.3059343>
- [9] Khurshid MR, Manzoor S, Sadiq T, Hussain L, Khan MS, Dutta AK. *Unveiling Diabetes Onset: Optimized XGBoost with Bayesian Optimization for Enhanced Prediction*. *PLoS One* [Internet]. 2025;20(1 January):1–29. Available from: <http://dx.doi.org/10.1371/journal.pone.0310218>
- [10] Wibowo A, Masruriyah AFN, Rahmawati S. *Refining Diabetes Diagnosis Models: The Impact of SMOTE on SVM, Logistic Regression, and Naïve Bayes for Imbalanced Datasets*. *J Electron Electromed Eng Med Informatics*. 2025;7(1):197–207. Available from: <https://doi.org/10.35882/jeemi.v7i1.596>
- [11] Fitriyah, L. (2024). *Implementasi Extreme Gradient Boosting (Xgboost) untuk Klasifikasi Risiko Diabetes Mellitus berdasarkan Riwayat Kesehatan guna meningkatkan Diagnosa Dini* (Doctoral dissertation, Universitas Islam Sultan Agung Semarang).
- [12] Laksono, H., Heriyanto, H., & Apriani, R. (2022). *Determinan Faktor Kejadian Komplikasi pada Penderita Diabetes Melitus di Kota Bengkulu Tahun 2021*. *Journal of Nursing and Public Health*, 10(1), 68-78.
- [13] Cahyaningtyas, U., & Werdiningsih, R. (2022). *Analisis Faktor Lama Penyembuhan Kaki Diabetes/Ulkus Diabetikum pada Pasien DM Tipe 2*. *Jurnal Media Administrasi*, 7(1), 28-39.
- [14] Istiono, W., Sutomo, A. H., Izhar, M. D., & Welembuntu, M. (2024). *Strategi Penerapan Patient Centered Care pada Pelayanan Kesehatan Primer*. UGM PRESS.
- [15] L. A. Puspita, R. (2023). *Hubungan Antara Perilaku Pencegahan Luka Kaki dan Gambaran Kaki Diabetik Pasien Diabetes Melitus* (Doctoral dissertation, Universitas Islam Sultan Agung Semarang).
- [16] Al-Hussein F, Tafakori L, Abdollahian M, Al-Shali K, Al-Hejin A. *Predicting Type 2 Diabetes Onset Age using Machine Learning: A Case Study in KSA*. *PLoS One* [Internet]. 2025;20(2 February). Available from: <https://doi.org/10.1371/journal.pone.0318484>
- [17] Zhou H, Xin Y, Li S. *A Diabetes Prediction Model based on Boruta Feature Selection and*

- Ensemble Learning*. BMC Bioinformatics [Internet]. 2023;24(1):1–34. Available from: <https://doi.org/10.1186/s12859-023-05300-5>
- [18] Lim H, Kim G, Choi JH. *Advancing Diabetes Prediction with a Progressive Self-Transfer Learning Framework for Discrete Time Series Data*. Sci Rep [Internet]. 2023;13(1):1–14. Available from: <https://doi.org/10.1038/s41598-023-48463-0>
- [19] Nilashi M, Abumalloh RA, Alyami S, Alghamdi A, Alrizq M. *A Combined Method for Diabetes Mellitus Diagnosis Using Deep Learning, Singular Value Decomposition, and Self-Organizing Map Approaches*. Diagnostics. 2023;13(10). Available from: <https://doi.org/10.3390/diagnostics13101821>
- [20] Setiawan, Zunan, et al. *Pengantar Sistem Informasi: Konsep Dasar dan Aplikasi Praktis*. PT. Sonpedia Publishing Indonesia, 2024..
- [21] Ramadhana, H. (2024). *Hubungan Tingkat Kepatuhan Minum Obat Anti Diabetes dengan Kadar Gula Darah pada Pasien Diabetes Melitus di Klinik Sakti Medika di Kota Nganjuk* (Doctoral Dissertation, STIKES Notokusumo Yogyakarta).
- [22] Ramadhani RN, Humairah I, Novida H, Qurnianingsih E. *Dyslipidemia is Associated with Coronary Heart Disease in Patients with Type 2 Diabetes Mellitus at a Tertiary Hospital in Surabaya, Indonesia*. 2025;06(1):24–30. Available from: <https://doi.org/10.20473/cimrj.v6i1.64560>
- [23] Hasanah U, Dewi NR, Ludiana L, Pakarti AT, Inayati A. *Analisis Faktor-Faktor Risiko Terjadinya Penyakit Ginjal Kronik pada Pasien Hemodialisis*. J Wacana Kesehat. 2023;8(2):96. Available from: <https://doi.org/10.52822/jwk.v8i2.531>
- [24] Norhalimi M, Siswa TAY. *Optimasi Seleksi Fitur Information Gain pada Algoritma Naïve Bayes dan K-Nearest Neighbor*. JISKA (Jurnal Inform Sunan Kalijaga). 2022;7(3):237–55. Available from: <https://doi.org/10.14421/jiska.2022.7.3.237-255>
- [25] Luthfiyyah IQ, Sari BN, Ridwan T. *Analisis Sentimen Mahasiswa terhadap Kurikulum Literasi Digital di Universitas Singaperbangsa Karawang menggunakan Naïve Bayes*. J Process. 2024;19(1):122–36. Available from: <https://doi.org/10.33998/processor.2024.19.1.1669>
- [26] Farhan NM, Setiaji B. *Indonesian Journal of Computer Science*. Indones J Comput Sci [Internet]. 2023;12(2):284–301. Available from: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
- [27] Aryani LN, Dantes GR, Ernanda Aryanto KY. *Pendekatan Zachman Framework untuk Perancangan Arsitektur Integrasi Data Sistem Remunerasi*. J Nas Pendidik Tek Inform. 2022;11(1):23. Available from: <https://doi.org/10.23887/janapati.v11i1.40181>
- [28] Palupi, L., Ihsanto, E., & Nugroho, F. (2023). *Analisis Validasi dan Evaluasi Model Deteksi Objek Varian Jahe Menggunakan Algoritma YOLOv5*. Journal of Information System Research (JOSH), 5(1), 234-241.
- [29] Swastika, R., Mukodimah, S., Susanto, F., Muslihudin, M., & Adab, S. I. P. (2023). *Implementasi Data Mining (Clustering, Association, Prediction, Estimation, Classification)*. Penerbit Adab.
- [30] Saracoglu BO. *Development Progress of Power Prediction Robot and Platform: Its World Level Very Long Term Prototyping Example*. J Energy Syst. 2022;6(2):253–89. Available from: <https://doi.org/10.30521/jes.1021838>