

**PERANCANGAN DAN IMPLEMENTASI SISTEM
DETEKSI DINI PENYAKIT *DIABETES*
MELLITUS MENGGUNAKAN
METODE *NAÏVE BAYES*
BERBASIS WEBSITE**

SKRIPSI

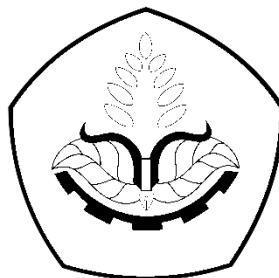


Oleh :
Brenda Rizky Tsania Gunawan
NIM. G41210047

**PROGRAM STUDI MANAJEMEN INFORMASI KESEHATAN
JURUSAN KESEHATAN
POLITEKNIK NEGERI JEMBER
2025**

**PERANCANGAN DAN IMPLEMENTASI SISTEM
DETEKSI DINI PENYAKIT *DIABETES*
MELLITUS MENGGUNAKAN
METODE *NAÏVE BAYES*
BERBASIS WEBSITE**

SKRIPSI



Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
Terapan Rekam Medis dan Informasi Kesehatan (S.Tr.RMIK) di
Program Studi Manajemen Informasi Kesehatan Jurusan Kesehatan

Oleh :
Brenda Rizky Tsania Gunawan
NIM. G41210047

**PROGRAM STUDI MANAJEMEN INFORMASI KESEHATAN
JURUSAN KESEHATAN
POLITEKNIK NEGERI JEMBER
2025**

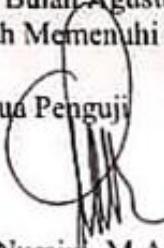
KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
POLITEKNIK NEGERI JEMBER
JURUSAN KESEHATAN

PERANCANGAN DAN IMPLEMENTASI SISTEM DETEKSI
DINI PENYAKIT *DIABETES MELLITUS* MENGGUNAKAN
METODE *NAÏVE BAYES* BERBASIS WEBSITE

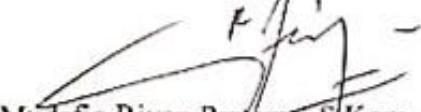
Brenda Rizky Tsania Gunawan (G41210047)

Telah Diuji pada Tanggal 21 Bulan Agustus Tahun 2025
dan Dinyatakan Telah Memenuhi Syarat

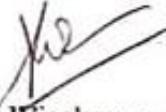
Ketua Penguji


dr. Novita Nuraini, M.A.R.S.
NIP. 19841107201012 2 002

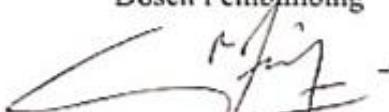
Sekertaris Penguji


Mudafiq Riyan Pratama, S.Kom., M.Kom.
NIP. 19890509201903 1 010

Anggota Penguji


Andri Permana Wicaksono, S.ST., M.T.
NIP. 19870305202321 1 011

Dosen Pembimbing


Mudafiq Riyan Pratama, S.Kom., M.Kom.
NIP. 19890509201903 1 010

Mengesahkan,
Ketua Jurusan Kesehatan


Ir. Rindiani, M.P.
NIP. 19680120199403 2 002

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Brenda Rizky Tsania Gunawan

NIM : G41210047

Menyatakan dengan sebenar-benarnya bahwa segala pernyataan dalam skripsi saya yang berjudul "Perancangan dan Implementasi Sistem Deteksi Dini Penyakit *Diabetes Mellitus* Menggunakan Metode *Naïve Bayes* Berbasis Website" merupakan gagasan dan hasil karya saya sendiri dengan arahan komisi pembimbing, dan belum pernah diajukan dalam bentuk apapun pada perguruan tinggi mana pun.

Semua data dan informasi yang digunakan telah dinyatakan secara jelas dan dapat diperiksa kebenarannya. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan dari penulis lain telah disebutkan dalam naskah dan dicantumkan dalam daftar pustaka di bagian akhir skripsi ini.

Jember, 15 Agustus 2025



Brenda Rizky Tsania Gunawan

NIM G41210047



**PERNYATAAN
PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Yang bertanda tangan di bawah ini, saya:

Nama : Brenda Rizky Tsania Gunawan
NIM : G41210047
Program Studi : Manajemen Informasi Kesehatan
Jurusan : Kesehatan

Demi pengembangan Ilmu Pengetahuan, saya menyetujui untuk memberikan kepada UPT. Perpustakaan Politeknik Negeri Jember, Hak Bebas Royalti *Non-Eksklusif* (*Non-Eksklusif Royalty Free Right*) atas Karya Ilmiah berupa Laporan Skripsi saya yang berjudul:

Perancangan dan Implementasi Sistem Deteksi Dini Penyakit *Diabetes Mellitus* Menggunakan Metode *Naïve Bayes* Berbasis Website

Dengan Hak Bebas Royalti *Non-Eksklusif* ini UPT. Perpustakaan Politeknik Negeri Jember berhak menyimpan, mengalih media atau format, mengelola dalam bentuk Pangkalan Data (*Database*), mendistribusikan karya dan menampilkan atau mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis atau penciptanya.

Saya bersedia untuk menanggung secara pribadi tanpa melibatkan pihak Politeknik Negeri Jember. Segala bentuk tuntutan hukum yang timbul atas Pelanggaran Hak Cipta dalam Karya ilmiah ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jember
Pada Tanggal : 15 Agustus 2025
Yang Menyatakan,



Nama : Brenda Rizky Tsania Gunawan
NIM : G41210047

MOTO

“Physically and mentally tired, but I still believe that,

فَإِنَّ مَعَ الْعُسْرِ يُسْرًا

In every hardship, there is relief.”

(QS. 94:6)

“Gonna be golden.”

(EJAE)

PERSEMBAHAN

Penulis menyadari dalam penyusunan Laporan Skripsi ini tidak lepas dari bantuan pihak yang memiliki peran besar, maka penulis mempersembahkan laporan skripsi ini kepada:

1. Allah SWT yang telah melimpahkan segala Rahmat, karunia serta hidayah Nya kepada penulis sehingga penyusunan laporan skripsi ini dapat terselesaikan dengan baik dan lancar.
2. Baginda Nabi Muhammad SAW serta keluarga dan para sahabat sebagai panutan umat muslim.
3. Kedua orang tua saya, Almarhum Bapak Hery Gunawan dan Ibu Nanik Wulandari yang senantiasa mendukung serta mendoakan saya dalam menempuh pendidikan hingga selesai. Terima kasih atas segala kasih sayang, pengorbanan, motivasi, dan dukungan baik secara moril maupun materil untuk kesuksesan penulis.
4. Kedua kakak laki-laki saya Mas Virza dan Mas Ronald yang selalu memberikan dukungan dan doa kepada saya dalam pembuatan Laporan Skripsi ini.
5. Dosen pembimbing saya, Bapak Mudafiq Riyam Pratama, S.Kom., M.Kom. yang senantiasa meluangkan waktunya untuk memberikan bimbingan, saran dan motivasi dalam menyelesaikan laporan skripsi ini.
6. Bapak Ibu dosen serta teknisi di program studi Manajemen Informasi Kesehatan yang selalu mengupayakan kegiatan perkuliahan dengan baik. Terima kasih atas ilmu pengetahuan, wawasan, serta saran juga nasihat yang bermanfaat.
7. 3 anomali yang menemani saya dari SMP dan teman-teman RJ yang selalu ada, terimakasih telah senantiasa memberikan motivasi, dukungan dan doa untuk saya dalam penyusunan laporan skripsi ini.
8. Teman-teman seperjuangan MIK angkatan 2021 yang sudah berbagi hari dan kisah kepada saya dalam menempuh pendidikan di Politeknik Negeri Jember.

9. Pimpinan dan seluruh petugas Rumah Sakit Citra Husada Jember yang telah membantu dalam memberikan data dan informasi yang dibutuhkan penulis terkait penyusunan laporan skripsi ini.
10. Almamater tercinta program studi Manajemen Informasi Kesehatan jurusan Kesehatan Politeknik Negeri Jember.

Perancangan dan Implementasi Sistem Deteksi Dini Penyakit *Diabetes Mellitus* Menggunakan Metode *Naïve Bayes* Berbasis Website
Mudafiq Riyam Pratama, S.Kom., M.Kom. (Pembimbing I)

Brenda Rizky Tsania Gunawan
Program Studi Manajemen Informasi Kesehatan
Jurusan Kesehatan

ABSTRAK

Diabetes Mellitus merupakan salah satu penyakit tidak menular dengan angka prevalensi yang terus meningkat setiap tahunnya, sehingga deteksi dini menjadi hal yang sangat penting untuk mencegah komplikasi lebih lanjut. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem deteksi dini *Diabetes Mellitus* berbasis website dengan menggunakan algoritma *Naïve Bayes*. Data yang digunakan berasal dari 600 rekam medis pasien, terdiri atas 300 pasien *Diabetes Mellitus* dan 300 pasien Hipertensi. Delapan variabel utama yang digunakan dalam penelitian ini adalah usia, jenis kelamin, indeks massa tubuh (IMT), tekanan darah, kadar glukosa darah acak (GDA), riwayat hipertensi, riwayat kardiovaskular, dan tingkat aktivitas fisik. Tahapan penelitian meliputi pengumpulan data, *pre-processing*, perancangan sistem dengan metode *Waterfall*, pemodelan menggunakan algoritma *Naïve Bayes*, hingga implementasi sistem. Hasil pengujian menunjukkan *accuracy* terbaik sebesar 73,81% dengan perbandingan data latih dan uji 65:35. Evaluasi melalui *confusion matrix* menghasilkan *accuracy* 77,5%, *precision* 100%, dan *recall* 72,7%. Sistem yang diimplementasikan telah melalui *blackbox testing* dan seluruh fungsionalitas utama berjalan sesuai harapan. Dengan demikian, sistem ini dapat menjadi alat bantu yang bermanfaat dalam mendukung deteksi dini *Diabetes Mellitus*, sekaligus meningkatkan kesadaran masyarakat akan pentingnya pemeriksaan kesehatan sejak dini.

Kata Kunci: *Diabetes Mellitus*, Rekam Medis, *Naïve Bayes*, Deteksi Dini, Sistem Informasi

*Design and Implementation of an Early Detection System for Diabetes Mellitus
Using the Naïve Bayes Method Based on a Website.
Mudafiq Riyana Pratama, S.Kom., M.Kom. (Chief Counselor).*

Brenda Rizky Tsania Gunawan
*Health Information Management Study Program
Departmen of Health*

ABSTRACT

Diabetes Mellitus is one of the leading non-communicable diseases with a steadily increasing prevalence, making early detection a crucial effort to prevent further complications. This study aims to design and implement an early detection system for Diabetes Mellitus based on medical records using the Naïve Bayes algorithm. The dataset consisted of 600 patient medical records, including 300 with Diabetes Mellitus and 300 with Hypertension. Eight main variables were employed in the classification process, namely age, gender, body mass index (BMI), blood pressure, random blood glucose (RBG), history of hypertension, history of cardiovascular disease, and level of physical activity. The research stages included data collection, preprocessing, system design using the Waterfall method, modeling with the Naïve Bayes algorithm, and system implementation. The experimental results showed the highest accuracy of 73.81% with a 65:35 ratio of training and testing data. Evaluation using the confusion matrix achieved an accuracy of 77.5%, precision of 100%, and recall of 72.7%. The implemented system was validated through blackbox testing, and all main functionalities performed as expected. Therefore, this system can serve as a valuable tool to support early detection of Diabetes Mellitus and raise public awareness regarding the importance of early health screening.

Keywords: *Diabetes Mellitus, Medical Records, Naïve Bayes, Early Detection, Information System*

RINGKASAN

Perancangan dan Implementasi Sistem Deteksi Dini Penyakit *Diabetes Mellitus* Menggunakan Metode *Naïve Bayes* Berbasis Website, Brenda Rizky Tsania Gunawan, NIM G41210047, D-IV Manajemen Informasi Kesehatan, Jurusan Kesehatan, Politeknik Negeri Jember, Mudafiq Riyam Pratama, S.Kom., M.Kom. (Pembimbing I).

Diabetes Mellitus merupakan salah satu penyakit tidak menular yang prevalensinya terus meningkat dari tahun ke tahun, baik di tingkat nasional maupun global. Penyakit ini ditandai dengan meningkatnya kadar glukosa darah yang tidak terkendali, sehingga dapat menimbulkan komplikasi serius seperti penyakit jantung, gagal ginjal, gangguan penglihatan, hingga amputasi. Kondisi tersebut menjadikan *Diabetes Mellitus* sebagai masalah kesehatan masyarakat yang membutuhkan perhatian serius, khususnya pada upaya deteksi dini. Deteksi dini dianggap penting karena semakin cepat risiko penyakit teridentifikasi, maka semakin besar peluang pasien untuk melakukan pencegahan dan pengendalian sebelum timbul komplikasi. Seiring perkembangan teknologi informasi di bidang kesehatan, sistem informasi berbasis rekam medis dapat dimanfaatkan untuk mendukung upaya deteksi dini. Data rekam medis mengandung informasi penting terkait faktor risiko pasien, seperti usia, jenis kelamin, indeks massa tubuh (IMT), tekanan darah, kadar glukosa darah, riwayat hipertensi, riwayat penyakit kardiovaskular, serta pola aktivitas fisik. Pemanfaatan data tersebut dengan pendekatan data mining memungkinkan terbangunnya sistem yang mampu memberikan prediksi risiko secara cepat, akurat, dan terukur.

Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem deteksi dini *Diabetes Mellitus* berbasis website dengan menggunakan algoritma *Naïve Bayes*. Algoritma ini dipilih karena kemampuannya dalam melakukan klasifikasi dengan memanfaatkan prinsip probabilistik, sederhana dalam implementasi, serta cukup efektif pada dataset dengan jumlah variabel terbatas. Dataset yang digunakan dalam penelitian ini berasal dari 600 rekam medis pasien Rumah Sakit Citra Husada Jember, yang terdiri atas 300 pasien dengan diagnosis

Diabetes Mellitus dan 300 pasien dengan diagnosis Hipertensi. Dari data tersebut, dipilih delapan variabel utama yang dianggap paling berpengaruh terhadap risiko *Diabetes Mellitus*. Tahapan penelitian dimulai dari pengumpulan data, *pre-processing* (meliputi pembersihan data, seleksi data, dan transformasi ke bentuk numerik), hingga perancangan sistem menggunakan model *Waterfall*. Pemodelan dilakukan dengan menguji beberapa komposisi data latih dan data uji. Hasil eksperimen menunjukkan bahwa *accuracy* terbaik diperoleh pada rasio data latih dan uji sebesar 65:35 dengan *accuracy* 73,81%. Selain itu, pengujian menggunakan *confusion matrix* memberikan hasil *accuracy* 77,5%, *precision* 100%, dan *recall* 72,7%. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan pasien dengan risiko *Diabetes Mellitus* secara cukup baik, meskipun masih terdapat beberapa keterbatasan pada sensitivitas deteksi kasus tertentu.

Sistem kemudian diimplementasikan dalam bentuk aplikasi berbasis website dengan antarmuka yang dirancang sesuai prinsip *user interface* (UI) dan *user experience* (UX). Fitur utama sistem meliputi menu *login*, *input* data pasien, proses deteksi dini, hasil prediksi, serta riwayat deteksi yang dapat diakses oleh pengguna. Proses pengkodean juga dilengkapi dengan validasi *input* dan normalisasi data untuk memastikan keakuratan hasil prediksi. Pengujian sistem dilakukan melalui *blackbox testing* yang difokuskan pada fungsionalitas utama. Hasil pengujian menunjukkan bahwa seluruh fungsi sistem berjalan sesuai spesifikasi yang dirancang, meliputi autentikasi pengguna, input faktor risiko, pemrosesan data dengan algoritma *Naïve Bayes*, hingga penyajian hasil prediksi kepada pengguna. Hal ini membuktikan bahwa sistem telah layak digunakan sebagai alat bantu deteksi dini risiko *Diabetes Mellitus*.

Secara keseluruhan, penelitian ini menghasilkan sebuah sistem deteksi dini *Diabetes Mellitus* berbasis website yang tidak hanya berperan sebagai alat pendukung keputusan bagi tenaga medis, tetapi juga dapat berfungsi sebagai media edukasi bagi masyarakat. Dengan adanya sistem ini, diharapkan masyarakat dapat lebih peduli terhadap kesehatan sejak dini, melakukan pemeriksaan secara berkala, serta mengurangi risiko komplikasi akibat *Diabetes Mellitus*.

PRAKATA

Segala puji syukur penulis haturkan kepada Allah SWT yang telah memberikan kelancaran, kemudahan, rahmat dan nikmat-Nya, yang begitu banyak sehingga laporan skripsi yang berjudul "Perancangan dan Implementasi Sistem Deteksi Dini Penyakit *Diabetes Mellitus* Menggunakan Metode *Naïve Bayes* Berbasis *Website*" dapat tersusun dengan baik.

Tulisan ini adalah laporan hasil penelitian yang dilaksanakan di Rumah Sakit Citra Husada Jember. Penyusunan laporan skripsi ini dimaksudkan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Terapan (S.Tr.RMIK) di Program Studi D-IV Manajemen Informasi Kesehatan Jurusan Kesehatan Politeknik Negeri Jember. Penyusunan laporan skripsi ini, penulis banyak mendapat bimbingan dan bantuan dari berbagai pihak sehingga dapat terselesaikan dengan baik. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Saiful Anwar, S.T.P., M.P. selaku Direktur Politeknik Negeri Jember.
2. Ir Rindiani, M.P selaku ketua Jurusan Kesehatan Politeknik Negeri Jember.
3. Dony Setiawan H.P., S.Kep., Ns., M.Kes selaku Ketua Program Studi DIV Manajemen Informasi Kesehatan Jurusan Kesehatan Politeknik Negeri Jember.
4. Mudafiq Riyam Pratama, S.Kom, M.Kom selaku dosen pembimbing saya.
5. Rekan-rekan dan semua pihak yang telah ikut serta membantu dalam pelaksanaan dan penulisan proposal skripsi ini.

Penulis menyadari, laporan skripsi ini masih jauh dari kata sempurna dan masih banyak kekurangan. Maka dari itu, kritik dan saran yang membangun akan penulis nantikan demi kesempurnaan proposal skripsi ini. Sehingga dapat penulis gunakan sebagai bahan evaluasi.

Jember, 15 Agustus 2025

DAFTAR ISI

	Halaman
HALAMAN SAMPUL.....	i
HALAMAN JUDUL	ii
HALAMAN PENGESAHAN SKRIPSI.....	iii
SURAT PERNYATAAN	iv
HALAMAN PUBLIKASI.....	Kesalahan!
Bookmark tidak ditentukan.	
MOTO	vi
PERSEMBAHAN.....	vii
ABSTRAK	ix
ABSTRACT	x
RINGKASAN	xi
PRAKATA	xiii
DAFTAR ISI.....	xiv
DAFTAR GAMBAR.....	xix
DAFTAR TABEL	xx
DAFTAR LAMPIRAN	xxi
BAB 1. PENDAHULUAN	2
1.1 Latar Belakang	2
1.2 Rumusan Masalah	7
1.3 Tujuan Penelitian	8
2.1.1 Tujuan Umum	8

2.1.2 Tujuan Khusus	8
1.4 Manfaat Penelitian	8
2.1.1 Bagi Rumah Sakit:	8
2.1.2 Bagi Politeknik Negeri Jember:	8
2.1.3 Bagi Peneliti:.....	9
 BAB 2. TINJAUAN PUSTAKA.....	 8
2.1 Peneliti Terdahulu	8
2.1.1 Sistem Pendukung Keputusan Berbasis Decision Tree Algorithm untuk.....	8
2.1.2 Perancangan Dan Pembuatan Sistem Deteksi Dini Penyakit Diabetes Melitus Di Puskesmas Sumbersari Jember Berbasis Website (Listyawati, 2024).....	8
2.2 State of The Art	10
2.3 Rumah Sakit	12
2.3.1 Pengertian Rumah Sakit.....	12
2.3.2 Tugas dan Fungsi Rumah Sakit	12
2.4 Rekam Medis	13
2.4.1 Pengertian Rekam Medis	13
2.4.2 Tujuan Rekam Medis.....	13
2.4.3 Manfaat Rekam Medis.....	13
2.5 Penyakit Diabetes Melitus (DM) Tipe 2	14
2.5.1 Definisi Diabetes Melitus (DM) Tipe 2.....	14
2.5.2 Faktor Risiko Diabetes Melitus (DM) Tipe 2	16
2.6 Data Mining	18
2.6.1 Definisi Data Mining	18
2.6.2 Tahapan Data Mining	18
2.6.3 Teknik Data Mining.....	20

2.7	Naïve Bayes	21
2.7.1	Langkah Perhitungan Metode Naïve Bayes.....	22
2.7.2	Proses Implementasi Algoritma Naïve Bayes	24
2.7.3	Kelebihan Metode Naïve Bayes	25
2.8	RapidMiner	26
2.9	Teknik Pemilihan Sampel Dalam Algoritma Naïve Bayes	26
2.10	Metode Waterfall.....	27
2.11	Confusion Matrix	29
2.12	Flowchart	31
2.13	Data Flow Diagram	32
2.14	Entity Relationshipship Diagram.....	33
2.15	Pemrograman Web	34
2.16	Hypertext Preprocessor	34
2.17	Blackbox Testing	35
2.18	Sistem Deteksi Dini.....	35
2.19	Kerangka Konsep	35
 BAB 3. METODE PENELITIAN.....		37
3.1	Jenis Penelitian	37
3.2	Waktu dan Tempat Penelitian	37
3.3	Variabel Penelitian	37
3.4	Objek Penelitian	38
3.5	Teknik Pengumpulan Data	38
3.6	Instrumen Penelitian	38
3.7	Definisi Operasional.....	39

3.8	Tahapan Penelitian.....	43
3.9	Pengujian Akurasi dengan Confusion Matrix	48
3.10	Map Menu Sistem.....	48
3.11	Contoh Penghitungan Klasifikasi Metode Naïve Bayes.....	51
BAB 4. HASIL DAN PEMBAHASAN		54
4.1	Mengumpulkan data dan pre-processing data untuk kebutuhan sistem deteksi dini penyakit Diabetes Mellitus (Communication).....	54
4.1.1	Pengumpulan Data	54
4.1.2	Pre-processing Data	55
4.1.3	Analisis Deskriptif Data Pasien	60
4.2	Membuat perencanaan proses pengembangan sistem deteksi dini penyakit Diabetes Mellitus (Planning).....	62
4.3	Menyusun dan menguji algoritma serta menganalisis dan mendesain sistem deteksi dini penyakit Diabetes Mellitus (Modelling).....	64
4.3.1	Pengujian Kinerja Algoritma Naïve Bayes.....	64
4.3.2	Analisis sistem	67
4.3.3	Desain Antarmuka Sistem	75
4.4	Melakukan pembangunan sistem deteksi dini penyakit Diabetes Mellitus (Construction)	80
4.4.1	Implementasi Antarmuka.....	80
4.4.2	Pengkodean Sistem	83
4.4.3	Pengujian Sistem Menggunakan Blackbox Testing	88
4.5	Menempatkan Sistem Deteksi Dini Penyakit Diabetes Mellitus	

pada Server Hosting (Deployment).....	89
BAB 5. KESIMPULAN DAN SARAN	91
5.1 Kesimpulan	91
5.2 Saran.....	92
DAFTAR PUSTAKA	94
LAMPIRAN.....	99

DAFTAR GAMBAR

	Halaman
Gambar 1. 1 Grafik 10 Besar Penyakit di RS Citra Husada 2024	4
Gambar 1. 2 Grafik Morbiditas Penyakit DM di Unit RI RS Citra Husada Jember 2020-2024	4
Gambar 2. 1 Metode Waterfall (Pressman et al., 2015).....	28
Gambar 2. 2 Kerangka Konsep	35
Gambar 4. 1 Tahapan penelitian	43
Gambar 4. 2 Map menu sistem	49
Gambar 4. 3 Gantt chart waterfall	63
Gambar 4. 4 Flowchart Sistem.....	68
Gambar 4. 5 Use case diagram.....	69
Gambar 4. 6 Use case diagram user	70
Gambar 4. 7 Activity diagram sistem deteksi dini	71
Gambar 4. 8 Entity-Relationship Diagram.....	73
Gambar 4. 9 Tampilan Home	75
Gambar 4. 10 Tampilan Pengisian Data Pasien	76
Gambar 4. 11 Tampilan Pemeriksaan Fisik	77
Gambar 4. 12 Tampilan Pengisian Riwayat Penyakit.....	78
Gambar 4. 13 Tampilan Hasil Deteksi Dini	79
Gambar 4. 14 Halaman login dokter dan admin	80
Gambar 4. 15 Halaman utama pasien	81
Gambar 4. 16 Halaman deteksi	81
Gambar 4. 17 Pop-up gagal.....	82
Gambar 4. 18 Halaman hasil deteksi.....	83
Gambar 4. 19 Proses validasi data	84
Gambar 4. 20 Proses normalisasi data	85
Gambar 4. 21 Proses probabilitas	87
Gambar 4. 22 Proses probabilitas posterior	88

DAFTAR TABEL

	Halaman
Tabel 2. 1 State Of The Art.....	10
Tabel 2. 2 Confusion Matrix	29
Tabel 2. 3 Simbol Flowchart.....	31
Tabel 2. 4 Simbol Data Flow Diagram	33
Tabel 2. 5 Simbol Entity Reltionshiop Diagram.....	33
Tabel 3. 1 Definisi Operasional	39
Tabel 3. 2 ata training.....	51
Tabel 3. 3 Probabilitas prior.....	51
Tabel 3. 4 Probabilitas Kondisional.....	52
Tabel 3. 5 Data testing	53
Tabel 3. 6 Probabilitas posterior	53
Tabel 3. 7 Hasil prediksi	53
Tabel 4. 1 Data Mentah Pasien Berisiko DM dan Tidak Berisiko DM ..	54
Tabel 4. 2 Data Kategorikal	56
Tabel 4. 3 Tabel kategorikal.....	57
Tabel 4. 4 Tabel numerikal	59
Tabel 4. 5 Tabel rekapitulasi data pasien	60
Tabel 4. 6 Hasil akurasi model.....	65
Tabel 4. 7 Perbandingan hasil klasifikasi dengan data aktual.....	65
Tabel 4. 8 Confusion matrix.....	66

DAFTAR LAMPIRAN

	Halaman
1 Lembar Observasi	100
2 Dokumentasi Studi Pendahuluan	100
3 Surat Rekomendasi Bankesbangpol.....	101
4 Surat Ijin Penelitian.....	102
5 Surat permohonan rekomendasi	103
6 Surat Keterangan Pengajuan Etik.....	104
7 Sertifikat turnitin	105
8 Dokumentasi penelitian.....	105
9 Surat balasan dari sakit.....	106

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Menurut *World Health Organization* (WHO), *Diabetes Melitus* (DM) adalah penyakit kronis yang terjadi ketika pankreas tidak memproduksi insulin dalam jumlah cukup atau ketika tubuh tidak dapat menggunakan insulin yang diproduksinya secara efektif. Insulin adalah hormon yang mengatur glukosa darah (WHO, 2024). Menurut *American Diabetes Association* atau (ADA, 2010), kekurangan insulin terjadi akibat sekresi insulin yang tidak mencukupi atau berkurangnya respon jaringan terhadap insulin di satu atau lebih titik dalam jalur kompleks kerja hormon. Penyakit *Diabetes Melitus* (DM) dapat menyebabkan komplikasi jangka panjang seperti retinopati, nefropati, neuropati perifer, dan neuropati otonom, serta meningkatkan risiko penyakit kardiovaskular, penyakit arteri perifer, dan penyakit serebrovaskular, hipertensi, dan gangguan metabolisme lipoprotein (ADA, 2010).

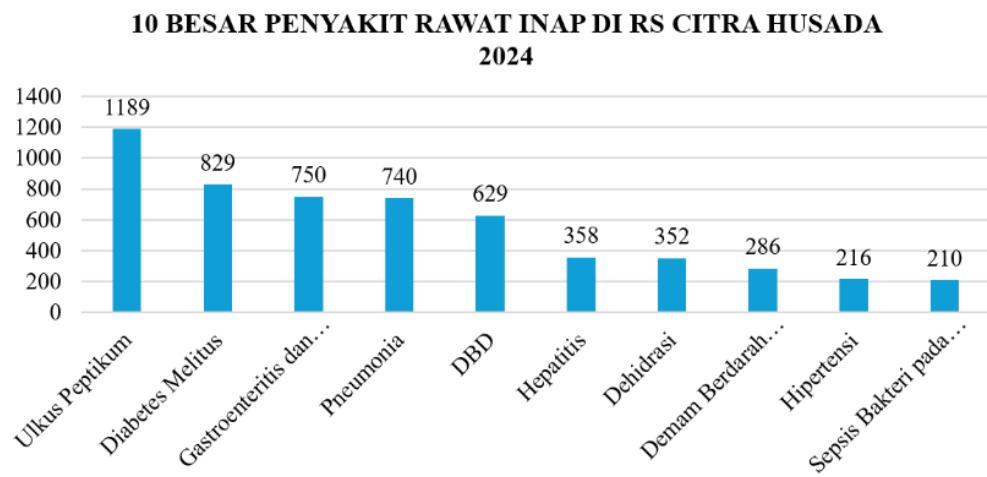
World Health Organization (WHO) menyatakan jumlah penderita *Diabetes Melitus* (DM) meningkat dari 200 juta pada tahun 1990 menjadi 830 juta pada tahun 2022. Prevalensi meningkat lebih cepat di negara-negara berpenghasilan rendah dan menengah daripada di negara-negara berpenghasilan tinggi (WHO, 2024). Meskipun data tersebut mencakup semua tipe DM, mayoritas kasus merupakan *Diabetes Mellitus* Tipe 2 (DM Tipe 2) yang erat kaitannya dengan pola hidup dan usia dewasa.

Organisasi *International Diabetes Federation* (IDF, 2021), memperkirakan jumlah penderita *Diabetes Melitus* (DM) di seluruh dunia sekitar 536,6 juta jiwa, dengan Indonesia menempati peringkat kelima dengan 19,5 juta jiwa berusia 20-79 tahun. Berdasarkan rentang usia tersebut, Sebagian besar kasus tergolong sebagai DM Tipe 2. Prevalensi penderita *Diabetes Melitus* (DM) yang terdiagnosis pada seluruh kelompok umur tertinggi terdapat di DKI Jakarta (2,6%), diikuti oleh D.I. Yogyakarta (2,4%), Sulawesi Utara (2,3%), Kalimantan Timur (2,3%) dan Jawa Timur (2,0%) (Riskesdas, 2018).

Data Riset Kesehatan Dasar tahun 2018 menunjukkan bahwa Provinsi Jawa Timur memiliki prevalensi Penyakit *Diabetes Melitus* (DM) sebesar 10,7%, menduduki peringkat ke-10 di Indonesia sebagai provinsi dengan jumlah penderita *Diabetes Melitus* (DM) tertinggi. Pada tahun 2018, prevalensi Penyakit *Diabetes Melitus* (DM) di Kabupaten Jember meningkat dibandingkan dengan Riset Kesehatan Dasar (RISKESDAS) 2013 berdasarkan pemeriksaan gula darah, yang menunjukkan Penyakit *Diabetes Melitus* (DM) meningkat dari 6,9% menjadi 8,5% (Duri et al., 2024). Data tersebut diperoleh dari hasil surveilans kasus penyakit tidak menular yang dikumpulkan dari rumah sakit di Kabupaten Jember, termasuk Rumah Sakit Citra Husada, salah satu fasilitas kesehatan yang berkontribusi dalam pencatatan dan pelaporan kasus *diabetes melitus* di wilayah tersebut. Peran aktif rumah sakit dalam mendukung pelaporan data kesehatan menjadi indikator penting dalam melihat tren penyakit yang berkembang di masyarakat.

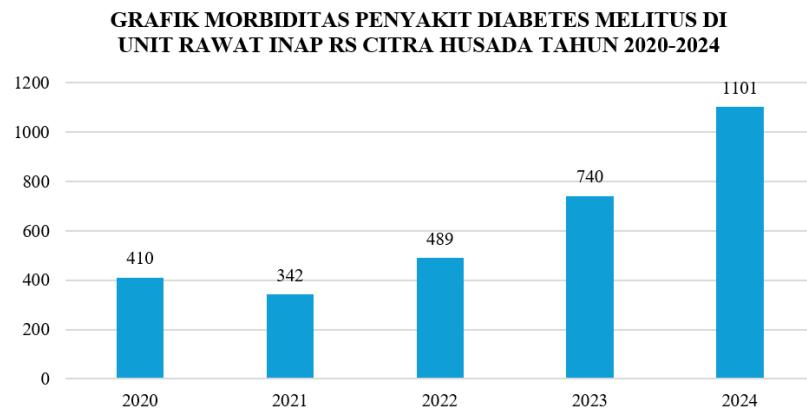
RS Citra Husada sebagai institusi pelayanan kesehatan yang cukup besar di wilayah Jember, dan menyimpan data yang representatif untuk dianalisis lebih lanjut. RS Citra Husada Jember merupakan rumah sakit tipe C yang beralamat di Jl. Teratai No. 22, Gebang Timur, Gebang, Kecamatan Patrang, Kabupaten Jember. Keberadaan rumah sakit ini tidak hanya penting dari sisi pelayanan kesehatan, tetapi juga sebagai sumber data sekunder yang relevan bagi penelitian di bidang rekam medis dan sistem informasi kesehatan.

Studi pendahuluan dilakukan untuk mengamati kondisi aktual di lapangan, dan melihat lebih jauh potensi penelitian yang bisa dikembangkan di rumah sakit ini. Berdasarkan hasil studi pendahuluan di RS Citra Husada Jember khususnya bagian rekam medis, ditemukan 10 besar penyakit pasien rawat inap di RS Citra Husada Jember tahun 2024 sebagai berikut:



Gambar 1. 1 Grafik 10 Besar Penyakit di RS Citra Husada 2024

Gambar 1.1 menunjukkan grafik 10 besar penyakit yang diperoleh dari laporan ringkasan morbiditas dan mortalitas rawat inap RS Citra Husada pada tahun 2024. Dari grafik di atas, kita dapat melihat bahwa *Diabetes Melitus* (DM) menempati peringkat kedua di antara 10 besar penyakit dengan 829 kasus. Selain itu penyakit *Diabetes Melitus* (DM) juga meningkat selama empat tahun terakhir.



Gambar 1. 2 Grafik Morbiditas Penyakit DM di Unit RI RS Citra Husada Jember 2020-2024

Gambar 1.2 menunjukkan tren kejadian *diabetes mellitus* (DM) di unit rawat inap Rumah Sakit Citra Husada dari tahun 2020 hingga 2024. Jumlah kasus berfluktuasi pada tahun 2021, terjadi penurunan dari 410 menjadi 342 kasus. Namun, dari tahun 2022 hingga 2024 terjadi peningkatan secara berturut-turut, dengan peningkatan terbesar terjadi pada tahun 2024. Jumlah

kasus meningkat dari 740 kasus pada tahun 2023 menjadi 1.101 kasus pada tahun 2024, atau meningkat 361 kasus. Secara keseluruhan, selama periode lima tahun, jumlah kasus meningkat 168,5%, yang menunjukkan tren peningkatan yang signifikan dalam tingkat *diabetes mellitus* dan membutuhkan lebih banyak perhatian dalam perencanaan intervensi dan manajemen layanan kesehatan. Peningkatan ini mencerminkan seriusnya *Diabetes Melitus* (DM) sebagai penyakit tidak menular yang fatal dan sering menimbulkan komplikasi berbahaya seperti penyakit jantung, gagal ginjal, dan gangguan saraf (Usman et al., 2021).

Menurut data terbaru yang diterbitkan oleh *International Diabetes Federation* (IDF) dalam *IDF Diabetes Atlas* edisi 2025, sekitar 589 juta orang dewasa (usia 20 hingga 79 tahun) hidup dengan menderita *diabetes mellitus* pada tahun 2024, yang mewakili 11,1% populasi orang dewasa di dunia (IDF, 2025a). Angka ini diperkirakan akan meningkat menjadi 853 juta pada tahun 2050 (IDF, 2025b). Hal ini sejalan dengan penelitian (Nugroho et al., 2023) yang menyatakan bahwa tingginya angka kematian akibat *Diabetes Melitus* (DM) merupakan masalah serius dan upaya pencegahan untuk menurunkan prevalensinya harus terus dilakukan. Pencegahan penyakit *Diabetes Melitus* (DM) dapat dilakukan dengan dua pendekatan, yaitu pencegahan primer dan sekunder. Salah satu bentuk intervensi yang dapat diterapkan adalah dengan memanfaatkan teknologi sistem informasi berbasis data rekam medis untuk mengembangkan model klasifikasi risiko *Diabetes Melitus* (DM).

Ada berbagai cara untuk mendeteksi *Diabetes Melitus* (DM), salah satunya dengan pendekatan Teknologi Informasi, seperti pengembangan sistem deteksi dini berbasis rekam medis. Data mining adalah cabang ilmu komputer yang mempelajari cara mengekstrak pengetahuan dan menemukan pola dari sejumlah besar data. Pengumpulan informasi data kemudian diproses untuk menghasilkan informasi yang lebih akurat atau menampilkan informasi sebaik mungkin (Fadillah et al., 2024). Metode data

mining dapat diterapkan untuk mengidentifikasi pola dan hubungan antara berbagai faktor risiko *Diabetes Melitus* (DM), yang membantu dalam proses klasifikasi penyakit. Klasifikasi adalah teknik analisis data yang mengekstrak model untuk menggambarkan kelas tertentu. Beberapa algoritma klasifikasi yang dapat digunakan adalah *K-Means*, *KNN*, *C.45*, *Naïve Bayes*, *Fuzzy*, dan lain-lain.

Algoritma *Naïve Bayes* merupakan algoritma sederhana yang menerapkan teorema Bayes dengan mengasumsikan bahwa semua fitur bersifat independen satu sama lain. Teorema Bayes merupakan dasar dari algoritma *Naïve Bayes* yang pertama kali digunakan oleh Thomas Bayes dalam teori probabilitas dan keputusan pada abad ke-18 (Al'Ariq et al., 2024). Algoritma *Naïve Bayes* dapat memprediksi keanggotaan kelas dalam suatu kelas menggunakan teori probabilitas. *Naïve Bayes* adalah algoritma pengelompokan atau klasifikasi dalam *Machine Learning*. Dibandingkan dengan algoritma lain, algoritma klasifikasi *Naïve Bayes* memiliki karakteristik kualitas yang baik, proses perhitungan yang sederhana, dan akurasi analisis yang tinggi (Fatmawati et al., 2023). Metode ini banyak digunakan di berbagai bidang, terutama bidang medis.

Penelitian (Afif, 2020) menunjukkan bahwa *Naïve Bayes* mengungguli algoritma lain dalam mengklasifikasikan penyakit *Diabetes Melitus* (DM) di Rumah Sakit Aisyiah. Keuntungan utama *Naïve Bayes* adalah mencapai tingkat akurasi tinggi sekitar 90,20% dibandingkan dengan algoritma lain. Selain itu, *Naïve Bayes* memiliki waktu komputasi yang rendah dan mudah diimplementasikan, menjadikannya pilihan yang efektif untuk aplikasi klasifikasi kasus *Diabetes Melitus* (DM). Sedangkan Penelitian (Rahmansyah, 2024) menunjukkan bahwa dibandingkan dengan K-Nearest Neighbor (KNN), *Naïve Bayes* memiliki performa yang lebih baik dalam hal akurasi. *Naïve Bayes* memiliki tingkat akurasi yang lebih tinggi sekitar 78,3582%, sedangkan KNN memiliki tingkat akurasi sekitar 77,9851%.

Keunggulan *Naïve Bayes* adalah bekerja dengan data pelatihan yang lebih sedikit dan membutuhkan waktu komputasi yang lebih sedikit.

Berdasarkan permasalahan yang ada, yaitu tingginya kasus *diabetes mellitus* yang terdapat pada RS Citra Husada Jember karena keterbatasan sistem monitoring yang cepat dan akurat, maka peneliti akan melakukan penelitian yang berjudul “Perancangan dan implementasi sistem deteksi dini penyakit *Diabetes Mellitus* dengan menggunakan metode *Naïve Bayes* berbasis website” Tujuan dari penelitian ini adalah untuk merancang sistem yang mampu mendeteksi risiko *diabetes mellitus* pada tahap awal dengan menggunakan data dari rekam medis pasien. Tujuan dari penelitian ini adalah untuk merancang sebuah sistem yang mampu mendeteksi risiko *diabetes mellitus* pada tahap awal dengan menggunakan data dari rekam medis pasien. Variabel yang digunakan dalam sistem adalah usia, jenis kelamin, tensi darah, indeks massa tubuh, kadar gula darah acak, riwayat hipertensi, riwayat kardiovaskular, kurangnya aktivitas yang merupakan faktor risiko *diabetes mellitus*. Dengan menggunakan metode *Naïve Bayes*, sistem akan secara otomatis dan efisien mengkategorikan risiko sehingga dapat secara teratur memantau kondisi kesehatan pasien dan memberikan peringatan dini ketika potensi risiko *diabetes mellitus* terdeteksi. Sistem ini akan membantu staf medis untuk membuat keputusan klinis dengan lebih cepat dan akurat.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, maka rumusan masalah dalam penelitian ini adalah bagaimana membuat sistem deteksi dini penyakit *Diabetes Mellitus* (DM) menggunakan metode *Naïve Bayes* berdasarkan data rekam medis dengan menggunakan metode pengembangan *waterfall* untuk menghasilkan sistem yang terorganisir dan dapat diimplementasikan dengan baik?

1.3 Tujuan Penelitian

2.1.1 Tujuan Umum

Merancang dan mengimplementasikan sistem deteksi dini penyakit *Diabetes Melitus* (DM) menggunakan metode *Naïve Bayes* berdasarkan data rekam medis.

2.1.2 Tujuan Khusus

- a. Mengumpulkan data dan *pre-processing* data untuk kebutuhan sistem deteksi dini penyakit *Diabetes Mellitus* (*Communication*)
- b. Membuat perencanaan proses pengembangan sistem deteksi dini penyakit *Diabetes Mellitus* (*Planning*)
- c. Menyusun dan menguji algoritma *Naïve Bayes* serta menganalisis dan mendesain sistem deteksi dini penyakit *Diabetes Mellitus* (*Modelling*)
- d. Melakukan pembangunan sistem deteksi dini penyakit *Diabetes Mellitus* (*Construction*)
- e. Menempatkan Sistem Deteksi Dini Penyakit *Diabetes Mellitus* pada *Server Hosting* (*Deployment*).

1.4 Manfaat Penelitian

2.1.1 Bagi Rumah Sakit:

- a. Hasil penelitian ini dapat mendeteksi risiko *Diabetes Melitus* (DM) pada pasien lebih awal dan lebih cepat, sehingga mereka dapat merawatnya dengan lebih efektif.
- b. Dapat membantu para tenaga medis merencanakan perawatan yang lebih baik dan sesuai kondisi pasien.
- c. Diharapkan dapat membantu meningkatkan kualitas pelayanan kesehatan secara keseluruhan dengan menyediakan data yang akurat dan dapat diandalkan.

2.1.2 Bagi Politeknik Negeri Jember:

- a. Memberikan kontribusi bagi pengembangan ilmu pengetahuan dan teknologi di bidang kesehatan, khususnya dalam penerapan algoritma *Naïve Bayes*.

- b. Menjadi referensi khususnya bagi mahasiswa D-IV Manajemen Informasi Kesehatan untuk mengembangkan penelitian di bidang yang sama.
- c. Membuka peluang untuk kolaborasi lebih lanjut antar instansi dalam pengembangan teknologi medis.

2.1.3 Bagi Peneliti:

- a. Membantu meningkatkan kemampuan dan keahlian peneliti di bidang data mining, *machine learning*, dan aplikasi teknologi dalam kesehatan.
- b. Meningkatkan kompetensi peneliti sebagai PMIK, aplikasi statistik kesehatan, epidemiologi dasar, dan biomedik terkait pemanfaatan data pelayanan dan program kesehatan untuk mengembangkan pelayanan kesehatan.

.

BAB 2. TINJAUAN PUSTAKA

2.1 Peneliti Terdahulu

2.1.1 Sistem Pendukung Keputusan Berbasis Decision Tree Algorithm untuk Prediksi Penyakit *Diabetes Melitus* (DM) (Karyadiputra & Setiawan, 2023)

Jurnal "Sistem Pendukung Keputusan Berbasis *Decision Tree Algorithm* untuk Prediksi Penyakit *Diabetes Melitus* (DM)" oleh Karyadiputra dan Setiawan (2023) memaparkan tentang pengembangan sistem pendukung keputusan untuk deteksi dini penyakit *Diabetes Melitus* (DM) menggunakan algoritma decision tree dengan menggunakan dataset dari *UCI Machine Learning Repository* yang terdiri dari 520 data pasien dan 17 variabel. Sistem dikembangkan dengan metodologi *waterfall* dan hasil pengujian menunjukkan akurasi prediksi sebesar 96,35%, lebih tinggi dibandingkan dengan *Naïve Bayes* (87,69%) dan KNN (89,62%). Validasi dilakukan dengan menggunakan *Confusion Matrix*, *Compare ROCs*, dan *Paired T-Test*, sehingga dapat disimpulkan bahwa decision tree merupakan metode yang paling efektif untuk memprediksi penyakit *Diabetes Melitus* (DM). Akan tetapi, sistem tersebut masih terbatas pada pengujian model dan belum diimplementasikan sebagai aplikasi berbasis web maupun digunakan pada data rekam medis sebenarnya dari rumah sakit di Indonesia.

2.1.2 Perancangan Dan Pembuatan Sistem Deteksi Dini Penyakit *Diabetes Melitus* Di Puskesmas Sumbersari Jember Berbasis Website (Listyawati, 2024)

Penelitian ini menguraikan tentang pengembangan sistem deteksi dini *Diabetes Melitus* (DM) berbasis website untuk Puskesmas Sumbersari Jember yang memiliki jumlah penderita DM terbanyak di Kabupaten Jember. Jumlah kunjungan pasien *Diabetes Melitus* (DM) meningkat signifikan dari 3.210 (2020) menjadi 9.432 (2022) karena rendahnya kesadaran masyarakat dan tidak berfungsinya pojok skrining. Sistem ini menggunakan metode Forward Chaining untuk penelusuran aturan dan Certainty Factor untuk memberi bobot kepastian diagnosis dan dirancang untuk memfasilitasi deteksi dini *Diabetes Melitus* (DM)

berdasarkan gejala pasien. Sistem ini dikembangkan menggunakan metodologi *waterfall* dan memiliki empat hak akses: dokter, petugas poli, petugas SIK dan

masyarakat. Data dikumpulkan melalui wawancara, observasi, dokumentasi, dan curah pendapat, dan dibuatlah basis data yang berisi 3 penyakit, 21 gejala, dan data pasien. Hasil penelitian menunjukkan bahwa sistem yang dikembangkan bermanfaat untuk mendeteksi dan memantau pasien *Diabetes Melitus* (DM), dan diharapkan dapat dikembangkan lebih lanjut dengan penambahan fungsionalitas untuk memantau sebaran pasien berdasarkan wilayah kerja puskesmas.

2.2 State of The Art

Perbedaan dan persamaan penelitian yang dilakukan oleh peneliti tersebut dengan penelitian yang dilakukan oleh peneliti lain adalah sebagai berikut:

Tabel 2. 1 *State Of The Art*

Materi	Karyadiputra & Listyawati (2024) Setiawan (2023)	Gunawan (2025)	
Bentuk Judul	Artikel Jurnal Sistem Pendukung Keputusan Berbasis <i>Decision Tree Algorithm</i> untuk Prediksi Penyakit Diabetes	Skripsi Perancangan Dan Pembuatan Sistem Deteksi Dini Penyakit <i>Diabetes Melitus</i> Di Puskesmas Sumbersari Jember Berbasis Website	Skripsi Perancangan dan Implementasi Sistem Deteksi Dini Penyakit <i>Diabetes Melitus</i> Menggunakan Metode <i>Naïve Bayes</i> Berbasis Data Rekam Medis
Metode	<i>Decision Tree</i>	<i>Forward Chaining</i> dan <i>Certainty Factor</i>	<i>Naïve Bayes</i>
Variabel Input	Status diabetes, Umur, Jenis kelamin, Kencing berlebih, Dehidrasi berlebih, BB turun secara signifikan, Merasa lemah, Lapar berlebih, Infeksi jamur ragi, Penglihatan kabur, Gatal-gatal, Cepat marah, Penyembuhan luka lambat, Otot melemah, Otot kaku, Rambut rontok, Obesitas	Usia <20 tahun, Usia >20 tahun, Sering buang air kecil, Sering lapar, Sering haus, Riwayat genetic, Penurunan berat badan <i>drastic</i> , Obesitas, Pusing, Mual, Muntah, Mudah lemas, Merasa Lelah, Mudah kantuk, Gatal-gatal, Kebas atau kesemutan, Luka sulit sembuh, Mudah infeksi, Pandangan kabur, Hipertensi, Nyeri otot	Usia, Jenis kelamin, Indeks Massa Tubuh, Kadar gula darah acak, Tensi darah, Riwayat hipertensi, Riwayat kardiovaskular, Kurangnya aktivitas

Materi	Karyadiputra & Setiawan (2023)	Listyawati (2024)	Gunawan (2025)
Variabel Output	a. Positif DM b. Negatif DM	a. Positif DM b. Negatif DM	a. Positif DM b. Negatif DM
Dataset	520 data pasien	255 data dikumpulkan melalui wawancara, observasi, dokumentasi, dan curah pendapat	600 dataset terdiri dari data penyakit <i>Diabetes Mellitus</i> dan Hipertensi
Akurasi Output	96,35% Analisis Akurasi	91% Sistem deteksi dini berbasis web	- Sistem deteksi dini berbasis web

Sumber : (Karyadiputra & Setiawan, 2023), (Listyawati, 2024)

Berdasarkan Tabel 2.1 di atas, dapat disimpulkan bahwa terdapat persamaan dan perbedaan antara penelitian ini dengan penelitian sebelumnya. Persamaannya dapat dilihat pada topik penelitian, yaitu terkait penerapan algoritma klasifikasi untuk mendeteksi penyakit, khususnya penyakit *Diabetes Mellitus* (DM). Perbedaannya terletak pada pendekatan yang digunakan, data yang dianalisis dan sistem yang dirancang, karena penelitian ini secara khusus berfokus pada perancangan dan implementasi sistem deteksi dini penyakit diabetes melitus berbasis data rekam medis dengan metode *Naïve Bayes*.

Penelitian yang dilakukan oleh Karyadiputra dan Setiawan (2023), metode *decision tree* digunakan untuk mendiagnosa penyakit *diabetes mellitus*, sedangkan Listyawati (2024) menggunakan pendekatan *forward chaining dan certainty factor*. Berbeda dengan kedua penelitian tersebut, Gunawan (2025) merancang dan mengembangkan sistem deteksi dini penyakit *diabetes mellitus* dengan menggunakan metode *Naïve Bayes* berbasis *website*. Tujuan dari penelitian ini adalah untuk mendeteksi dini risiko *diabetes mellitus* dengan cara yang praktis menggunakan sistem yang terintegrasi. Kelebihan dari penelitian ini terletak pada penggunaan data set yang lebih besar dan penggunaan faktor risiko daripada gejala. *Diabetes Mellitus* (DM) adalah penyakit yang berkembang secara progresif dan pada awalnya tidak menunjukkan gejala. Gejala-gejala seperti sering buang air kecil, haus dan lapar biasanya tidak disadari hingga penyakit ini sudah parah, yang menjelaskan mengapa DM sering disebut sebagai ‘*the silent killer*’ (Isnaini & Ratnasari, 2018).

2.3 Rumah Sakit

2.3.1 Pengertian Rumah Sakit

Menurut UU No. 17 tahun 2023 tentang Kesehatan, rumah sakit adalah fasilitas pelayanan kesehatan yang menyelenggarakan Pelayanan Kesehatan perseorangan secara paripurna melalui pelayanan kesehatan promotif, preventif, kuratif, rehabilitatif, dan/ atau paliatif dengan menyediakan pelayanan rawat inap, rawat jalan, dan gawat darurat. Rumah sakit biasanya memiliki berbagai spesialisasi medis dan dilengkapi dengan peralatan medis terkini untuk menangani berbagai kondisi kesehatan. Rumah sakit bertujuan untuk menyediakan pelayanan kesehatan yang aman, berkualitas, dan terjangkau bagi masyarakat.

2.3.2 Tugas dan Fungsi Rumah Sakit

Tugas rumah sakit menurut UU No. 17 tahun 2023 tentang Kesehatan adalah sebagai berikut:

- a. Memberikan layanan kesehatan yang komprehensif, meliputi pelayanan promotif, preventif, kuratif, rehabilitatif, dan paliatif;
- b. Menangani kondisi pasien dalam situasi darurat;
- c. Memberikan pendidikan dan pelatihan bagi tenaga kesehatan;
- d. Melaksanakan penelitian untuk pengembangan ilmu dan layanan kesehatan.

Adapun fungsi rumah sakit menurut UU No. 17 tahun 2023 tentang Kesehatan adalah sebagai berikut:

- a. Sebagai pusat rujukan kesehatan bagi masyarakat dan fasilitas kesehatan lainnya.
- b. Mengintegrasikan sistem informasi kesehatan untuk meningkatkan efisiensi pelayanan.
- c. Berperan untuk meningkatkan kesehatan masyarakat melalui program kesehatan.
- d. Meningkatkan akses ke layanan kesehatan, termasuk penggunaan teknologi medis mutakhir.

2.4 Rekam Medis

2.4.1 Pengertian Rekam Medis

Rekam medis adalah dokumen yang memuat catatan vital pasien, data mengenai identitas pasien, riwayat kesehatan, hasil pemeriksaan fisik, catatan seluruh kegiatan tenaga kesehatan terhadap pasien yang diperiksa dari waktu ke waktu dan yang menerima pelayanan medis di sarana pelayanan kesehatan melalui pengobatan rawat jalan atau rawat inap (Prasetyo et al., 2023). Berdasarkan Peraturan Menteri Kesehatan Republik Indonesia No. 24 Tahun 2022 tentang Rekam Medis, rekam medis adalah dokumen yang memuat data identitas pasien, pemeriksaan, pengobatan, tindakan, dan pelayanan lain yang diberikan kepada pasien. Rekam medis ini dapat dalam format elektronik dan harus dikelola oleh fasilitas pelayanan kesehatan di Indonesia.

2.4.2 Tujuan Rekam Medis

Berdasarkan Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022, tujuan rekam medis adalah:

- a. Meningkatkan mutu pelayanan kesehatan;
- b. Memberikan kepastian hukum dalam penyelenggaraan dan pengelolaan rekam medis;
- c. Menjamin keamanan, kerahasiaan, keutuhan, dan ketersediaan data rekam medis; dan
- d. Mewujudkan penyelenggaraan dan pengelolaan rekam medis yang terintegrasi secara digital.

2.4.3 Manfaat Rekam Medis

Menurut Yustina EW (2012) yang dikutip dalam penelitian (Amran et al., 2022), dalam pelaksanaannya rekam medis disusun dengan tujuan untuk menciptakan ketertiban administrasi dengan tujuan untuk meningkatkan pelayanan medis di rumah sakit yang didukung oleh sistem pengelolaan rekam medis yang baik dan akurat. Sejalan dengan tujuan rekam medis, rekam medis juga dibuat untuk memberikan informasi yang lengkap dan akurat serta disiapkan untuk diberikan dalam jangka waktu tertentu dalam pelaksanaan pelayanan medis.

Rekam medis, sebagai catatan yang akurat dan lengkap, menggambarkan status kesehatan pasien, termasuk penyakit masa lalu dan sekarang serta pengobatannya.

2.5 Penyakit *Diabetes Melitus* (DM) Tipe 2

2.5.1 Definisi *Diabetes Melitus* (DM) Tipe 2

Diabetes Melitus (DM) merupakan penyakit yang disebabkan oleh penumpukan glukosa dalam darah, yang terjadi ketika tubuh tidak memproduksi cukup insulin atau tidak mampu menggunakan insulin dengan baik, dan biasanya ditandai dengan buang air kecil terus-menerus (dalam jumlah banyak) dengan rasa manis (kencing manis) (Bilous & Donelly, 2014). Resistensi insulin, akibat kerusakan sel-β pankreas dan kurangnya kemampuan insulin untuk merangsang penggunaan glukosa atau kurangnya respons sel target, seperti otot, jaringan, dan hati, terhadap kadar insulin fisiologis, juga merupakan penyebab utama *Diabetes Melitus* (DM). *Diabetes Melitus* (DM) merupakan salah satu PTM paling signifikan secara global dan penyebab utama menurunnya kualitas hidup (Tamornpark et al., 2022).

Klasifikasi etiologi *Diabetes Melitus* (DM) menurut *American Diabetes Association* tahun 2017 dibagi menjadi empat tipe (ADA, 2017).

1) *Diabetes Melitus* (DM) tipe I atau *Insulin Dependent Diabetes Mellitus* (IDDM)

Diabetes Melitus (DM) tipe I terjadi karena rusaknya sel-sel beta di pankreas akibat faktor autoimun. Pada *Diabetes Melitus* (DM) tipe ini, sekresi insulin sedikit atau tidak ada sama sekali. Hal ini dapat diketahui dari kadar protein C-peptida yang rendah atau tidak terdeteksi sama sekali. Gejala klinis pertama penyakit ini adalah ketoasidosis. *Diabetes Melitus* (DM) tipe 1 disebabkan oleh kerusakan sistem imun yang disebabkan oleh infeksi virus atau reaksi autoimun yang merusak sel-sel penghasil insulin, yaitu sel-sel beta di seluruh pankreas. Oleh karena itu, pada *Diabetes Melitus* (DM) tipe I, pankreas tidak dapat memproduksi insulin. Penderita *Diabetes Melitus* (DM) harus menerima insulin melalui suntikan ke dalam tubuhnya agar dapat bertahan hidup. Jika insulin tidak diberikan, pasien akan jatuh pingsan, yang juga dikenal sebagai koma ketoasidosis atau koma diabetik (Nurrahmani, 2012).

2) *Diabetes Melitus* (DM) tipe II atau *Insulin Non-dependent Diabetes Mellitus* (NIDDM)

Pasien dengan *Diabetes Melitus* (DM) tipe ini mengalami hiperinsulinemia, tetapi karena resistensi insulin, insulin tidak dapat mengangkut glukosa ke jaringan. Resistensi insulin adalah berkurangnya kemampuan insulin untuk merangsang penyerapan glukosa oleh jaringan perifer dan menghambat produksi glukosa oleh hati. Retensi insulin (kadar insulin darah masih dianggap tinggi, sehingga reseptor insulin tidak aktif) menyebabkan kekurangan insulin relatif. Hal ini dapat mengakibatkan menurunnya sekresi insulin lain dan desensitivasi β pankreas terhadap keberadaan glukosa (Maulana, 2009). *Diabetes Melitus* (DM) tipe II disebabkan oleh insufisiensi relatif sel β pankreas dan resistensi insulin. Resistensi insulin adalah berkurangnya kemampuan insulin untuk merangsang penyerapan glukosa oleh jaringan perifer dan menghambat produksi glukosa oleh hati. Ketidakmampuan sel β pankreas untuk sepenuhnya mengimbangi resistensi insulin ini mengakibatkan defisiensi insulin relatif. Ketidakmampuan ini terlihat dari berkurangnya sekresi insulin pada rangsangan glukosa maupun pada rangsangan glukosa bersama bahan perangsang sekresi insulin lain (Sulistyowati, 2016). Salah satu penyebab *Diabetes Melitus* (DM) terutama DM tipe 2, adalah keadaan stres oksidatif yang menyebabkan resistensi insulin dari sel β pankreas. Selain itu, hiperglikemia juga terlibat dalam proses pembentukan radikal bebas. Hiperglikemia menyebabkan auto-oksidasi glukosa, glikasi protein, dan aktivasi jalur metabolisme poliol, yang mempercepat pembentukan senyawa oksigen reaktif. Oksidasi lipid yang berlebihan dapat membentuk senyawa radikal, yang membutuhkan senyawa antioksidan untuk menguranginya (Triandita et al., 2016).

3) *Diabetes Melitus* (DM) Tipe Lain.

Diabetes Melitus (DM) tipe ini disebabkan oleh gangguan metabolisme yang ditandai dengan kadar glukosa darah yang tinggi akibat faktor genetik yang memengaruhi fungsi sel beta, penyakit pankreas eksokrin, gangguan endokrin-metabolik lainnya, dan infeksi virus. *Diabetes Melitus* (DM) tipe ini juga dapat disebabkan oleh obat-obatan atau bahan kimia (contohnya, dalam pengobatan HIV/AIDS atau setelah transplantasi organ) (ADA, 2017).

4) *Diabetes Melitus (DM) Gestasional.*

Diabetes Melitus (DM) tipe ini berkembang selama kehamilan, dan gangguan toleransi glukosa biasanya pertama kali terdeteksi selama trimester kedua dan ketiga kehamilan. *Diabetes Melitus (DM)* gestasional dikaitkan dengan meningkatnya komplikasi perinatal. Wanita dengan *Diabetes Melitus (DM)* gestasional memiliki risiko lebih tinggi terkena *Diabetes Melitus (DM)* yang berlangsung selama 5 hingga 10 tahun setelah melahirkan (Lanywati, 2001).

2.5.2 Faktor Risiko *Diabetes Melitus (DM)* Tipe 2

Faktor risiko merupakan kondisi, kebiasaan, atau karakteristik tertentu yang meningkatkan kemungkinan seseorang mengalami suatu penyakit. Dalam konteks *Diabetes Melitus (DM)*, faktor risiko dikategorikan menjadi dua, yaitu faktor risiko yang tidak dapat dimodifikasi dan yang dapat dimodifikasi. Pemahaman terhadap kedua kelompok faktor risiko ini penting dalam upaya promotif dan preventif terhadap penyakit DM Tipe 2 (P2PTM, 2008).

a. Faktor Risiko yang Tidak Dapat Dimodifikasi

Faktor risiko ini bersifat tetap dan tidak bisa diubah oleh individu. Deteksi terhadap faktor-faktor ini biasanya dilakukan melalui tanya jawab atau pengisian riwayat kesehatan. Beberapa faktor yang termasuk dalam kategori ini antara lain:

1. Riwayat keluarga dengan DM

Seseorang yang memiliki anggota keluarga (ayah, ibu, atau saudara kandung) yang menderita DM memiliki risiko lebih tinggi untuk mengalami penyakit serupa. Faktor genetik berperan besar dalam memengaruhi kerentanan terhadap gangguan metabolisme glukosa.

2. Umur

Risiko menderita prediabetes maupun DM meningkat seiring bertambahnya usia. Proses penuaan menyebabkan penurunan sensitivitas insulin dan fungsi sel beta pankreas, sehingga individu usia lanjut lebih rentan mengalami gangguan metabolisme.

3. Riwayat pernah menderita diabetes gestasional

Wanita yang pernah mengalami diabetes selama masa kehamilan (diabetes gestasional) memiliki kemungkinan lebih besar untuk mengalami

DM tipe 2 di kemudian hari, terutama jika tidak melakukan perubahan gaya hidup setelah kehamilan.

4. Riwayat berat badan lahir rendah (BBLR)

Berat badan lahir kurang dari 2500 gram merupakan indikator pertumbuhan intrauterin yang tidak optimal, dan telah terbukti berhubungan dengan peningkatan risiko resistensi insulin serta gangguan metabolisme glukosa di usia dewasa.

b. Faktor Risiko yang Dapat Dimodifikasi

Faktor risiko ini berkaitan dengan gaya hidup dan kondisi kesehatan yang dapat diubah atau dikendalikan melalui perubahan perilaku dan intervensi medis. Beberapa faktor tersebut antara lain:

1. Berat badan lebih

Kelebihan berat badan ($BB >120\%$ dari BB ideal atau indeks massa tubuh/IMT $>23 \text{ kg/m}^2$) serta lingkar pinggang berlebih ($\geq 90 \text{ cm}$ untuk pria dan $\geq 80 \text{ cm}$ untuk wanita) meningkatkan risiko resistensi insulin. Lemak visceral (lemak perut) sangat berperan dalam perkembangan DM tipe 2 karena mengganggu metabolisme glukosa.

2. Kurangnya aktivitas fisik

Gaya hidup sedentari atau kurang berolahraga menyebabkan penurunan sensitivitas insulin dan peningkatan kadar glukosa darah. Aktivitas fisik rutin membantu meningkatkan penggunaan glukosa oleh otot dan mengontrol berat badan.

3. Hipertensi

Tekanan darah $\geq 140/90 \text{ mmHg}$ berhubungan dengan peningkatan risiko DM. Hipertensi dan DM berbagi jalur patofisiologi yang sama, termasuk resistensi insulin, peradangan kronis, dan disfungsi endotel.

4. Dislipidemia

Kondisi ini ditandai dengan ketidakseimbangan kadar lipid darah seperti:

- Kolesterol HDL rendah ($\leq 35 \text{ mg/dl}$)
- Triglicerida tinggi ($\geq 250 \text{ mg/dl}$)

Dislipidemia mempercepat terjadinya resistensi insulin dan gangguan metabolisme lemak yang berkontribusi terhadap DM tipe 2.

5. Riwayat penyakit kardiovaskular

Penderita penyakit jantung atau pembuluh darah sering mengalami gangguan metabolisme glukosa. DM dan penyakit kardiovaskular saling berkaitan dan dapat memperburuk kondisi satu sama lain.

6. Pola diet yang tidak sehat

Konsumsi makanan tinggi kalori, tinggi gula, tinggi lemak jenuh, serta rendah serat merupakan faktor penting yang meningkatkan risiko DM. Diet tidak seimbang memengaruhi metabolisme tubuh, berat badan, dan sensitivitas insulin secara keseluruhan.

2.6 Data Mining

2.6.1 Definisi Data Mining

Data mining merupakan proses penggalian pola dan pengetahuan dari sekumpulan data dalam jumlah besar (Arhami & Nasir, 2020). Proses ini bertujuan untuk mengekstrak informasi yang bernilai dari data yang tersedia, sehingga dapat digunakan untuk mendukung pengambilan keputusan. Data mining juga dikenal sebagai bagian dari Knowledge Discovery in Database (KDD), di mana informasi tersembunyi yang bersifat implisit dapat diidentifikasi dan diolah menjadi wawasan yang dapat dimanfaatkan untuk berbagai keperluan, termasuk dalam bidang kesehatan seperti deteksi dini penyakit *Diabetes Melitus* (DM) berbasis data rekam medis (Ali Ikhwan & Aslami, 2024).

2.6.2 Tahapan Data Mining

Dalam implementasi data mining, terdapat beberapa tahap utama yang harus dilakukan agar proses analisis berjalan secara sistematis dan menghasilkan informasi yang valid. Tahapan tersebut meliputi:

a. *Data Cleaning* (Pembersihan Data)

Proses ini bertujuan untuk menyaring dan menghapus data yang tidak relevan atau mengalami inkonsistensi. Dalam pengolahan data rekam medis, kemungkinan terdapat data yang tidak lengkap atau tidak valid. Oleh karena itu, data yang

memiliki kesalahan atau ketidaksesuaian harus diperbaiki atau dihapus agar tidak mempengaruhi hasil analisis.

b. *Data Integration* (Integrasi Data)

Data yang digunakan dalam penelitian dapat berasal dari berbagai sumber. Oleh karena itu, tahap integrasi dilakukan untuk menggabungkan data dari berbagai sumber tersebut menjadi satu kesatuan yang dapat dianalisis secara lebih komprehensif.

c. *Data Selection* (Seleksi Data)

Tidak semua data yang tersedia akan digunakan dalam analisis. Oleh sebab itu, hanya data yang relevan dan sesuai dengan tujuan penelitian yang akan dipilih. Sebagai contoh, dalam sistem deteksi dini *Diabetes Melitus* (DM), data yang digunakan dapat mencakup informasi seperti usia, jenis kelamin, tensi darah, indeks massa tubuh, kadar gula darah acak, riwayat hipertensi, riwayat kardiovaskular, kurangnya aktivitas sedangkan informasi lain yang tidak berhubungan tidak dimasukkan dalam analisis.

d. *Data Transformation* (Transformasi Data)

Proses ini dilakukan untuk mengonversi data ke dalam format yang lebih sesuai untuk analisis lebih lanjut. Sebagai contoh, jika data rekam medis memiliki variabel dalam bentuk angka numerik yang perlu dikategorikan, maka data tersebut akan dikelompokkan ke dalam beberapa interval agar lebih mudah diolah oleh algoritma data mining.

e. *Mining Process* (Proses Penambangan Data)

Tahapan ini merupakan inti dari data mining, di mana metode analisis diterapkan untuk menemukan pola atau hubungan dalam data. Dalam konteks deteksi dini penyakit *Diabetes Melitus* (DM), algoritma *Naïve Bayes* digunakan untuk mengklasifikasikan data pasien berdasarkan faktor-faktor risiko yang ada.

f. *Pattern Evaluation* (Evaluasi Pola)

Setelah pola ditemukan, langkah selanjutnya adalah mengevaluasi hasil tersebut guna memastikan bahwa pola yang dihasilkan benar-benar sesuai dengan tujuan penelitian. Jika hasil yang diperoleh kurang akurat, maka dapat dilakukan penyesuaian pada parameter analisis atau menggunakan metode lain yang lebih sesuai.

g. Knowledge Presentation (Penyajian Pengetahuan)

Tahap akhir dalam data mining adalah menyajikan hasil analisis dalam bentuk yang mudah dipahami, seperti tabel, grafik, atau laporan. Penyajian ini bertujuan agar informasi yang diperoleh dapat digunakan secara efektif oleh pengguna, seperti dokter atau tenaga medis dalam proses deteksi dini penyakit *Diabetes Melitus* (DM).

2.6.3 Teknik Data Mining

Dalam data mining, terdapat berbagai teknik yang digunakan untuk mengekstrak informasi dari kumpulan data besar. Beberapa teknik utama yang sering digunakan dalam analisis kesehatan meliputi:

a. Klasifikasi

Teknik klasifikasi digunakan untuk memprediksi kelas dari suatu data berdasarkan pola yang telah dipelajari dari data sebelumnya. Dalam penelitian ini, metode *Naïve Bayes* digunakan sebagai teknik klasifikasi utama untuk menentukan apakah seorang pasien memiliki risiko tinggi atau rendah terkena *Diabetes Melitus* (DM) berdasarkan data rekam medisnya. Algoritma klasifikasi lainnya yang sering digunakan dalam bidang kesehatan meliputi *Support Vector Machine* (SVM), ID3, dan Random Forest.

b. Klasterisasi

Klasterisasi adalah metode yang digunakan untuk mengelompokkan data tanpa menggunakan label kelas sebelumnya. Teknik ini sering digunakan untuk menemukan kelompok pasien dengan karakteristik serupa berdasarkan gejala atau

riwayat kesehatan mereka. Beberapa metode klasterisasi yang umum digunakan antara lain K-Means, Fuzzy C-Means, dan K-Medoids.

c. Asosiasi

Teknik asosiasi bertujuan untuk menemukan hubungan antara satu variabel dengan variabel lainnya dalam suatu dataset. Misalnya, dalam bidang medis, teknik ini dapat digunakan untuk mengidentifikasi pola keterkaitan antara gaya hidup pasien dengan risiko terkena *Diabetes Melitus* (DM). Algoritma asosiasi yang sering digunakan mencakup Apriori, Chi-Square, dan FP-Growth.

d. Regresi

Teknik regresi digunakan untuk memprediksi nilai numerik berdasarkan hubungan antara variabel input dan output. Dalam sistem deteksi dini *Diabetes Melitus* (DM), teknik regresi dapat digunakan untuk memperkirakan kadar gula darah berdasarkan usia, indeks massa tubuh, dan riwayat kesehatan pasien. Algoritma regresi yang umum digunakan meliputi *Linear Regression*, *Logistic Regression*, dan *Support Vector Regression* (Muslim et al., 2019).

2.7 *Naïve Bayes*

Metode *Naïve Bayes* merupakan salah satu metode dalam data mining yang digunakan untuk memprediksi probabilitas suatu kejadian berdasarkan pengalaman sebelumnya (Rozi, 2023). Inti dari metode ini adalah menentukan kemungkinan kelas suatu data berdasarkan probabilitas yang dihitung dari data sebelumnya. Metode ini termasuk dalam kategori klasifikasi statistik, di mana tujuan utamanya adalah menentukan kemungkinan kelas suatu data berdasarkan probabilitas yang diperoleh dari data sebelumnya. Dalam konteks klasifikasi, *Naïve Bayesian Classifier* adalah pendekatan yang didasarkan pada teorema Bayes, dengan asumsi bahwa setiap variabel atau atribut dalam data bersifat independen satu sama lain (Abdussomad et al., 2024). Oleh karena itu, metode ini disebut sebagai "Naïve" karena menganggap bahwa tidak ada hubungan antar atribut, meskipun dalam kenyataan, atribut dalam dataset sering kali memiliki keterkaitan.

Metode *Naïve Bayes Classifier* bekerja berdasarkan teorema Bayes, yang secara umum memiliki bentuk sebagai berikut (Dahri, Agus, & Khairina, 2016):

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)}$$

Keterangan:

X = Data yang akan diklasifikasikan ke dalam suatu kelas tertentu.

C = Hipotesis bahwa data X termasuk dalam suatu kelas tertentu.

P(C) = Probabilitas awal dari kelas C (probabilitas prior).

P(X) = Probabilitas keseluruhan data X dalam dataset (probabilitas evidence).

P(C|X) = Probabilitas bahwa X termasuk dalam kelas C berdasarkan data yang tersedia (probabilitas posterior).

P(X|C) = Probabilitas data X terjadi dalam kondisi kelas C (probabilitas likelihood).

Dengan menggunakan teorema Bayes, klasifikasi data dilakukan berdasarkan nilai posterior, yang dihitung menggunakan persamaan:

$$\text{Posterior} \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

Di dalam metode *Naïve Bayes*, klasifikasi dilakukan dengan membandingkan nilai probabilitas dari setiap kelas yang mungkin terjadi. Kelas dengan nilai probabilitas tertinggi akan dipilih sebagai hasil klasifikasi.

2.7.1 Langkah Perhitungan Metode *Naïve Bayes*

Untuk mendapatkan hasil klasifikasi yang optimal menggunakan metode *Naïve Bayes*, terdapat beberapa langkah yang harus dilakukan, yaitu:

1. Menentukan Probabilitas Prior (P(C))
 - a. Probabilitas prior dihitung berdasarkan rasio jumlah data dalam setiap kelas dibandingkan dengan total jumlah data dalam dataset.

Rumus yang digunakan adalah:

$$P(C) = \frac{X}{A}$$

Keterangan:

$P(C)$ = Probabilitas prior dari kelas C

X = Jumlah data dalam kelas tertentu

A = Jumlah total data dalam dataset

2. Menentukan Probabilitas Likelihood ($P(X|C)$)

- a. Probabilitas likelihood dihitung berdasarkan jumlah kejadian suatu atribut dalam kelas tertentu dibandingkan dengan total jumlah atribut dalam kelas tersebut.

b. Rumus yang digunakan adalah:

$$P(X/C) = \frac{F}{B}$$

Keterangan:

$P(X|C)$ = Probabilitas likelihood suatu atribut X dalam kelas C

F = Jumlah data fitur dalam kelas tertentu

B = Jumlah total fitur dalam kelas tersebut

3. Menghitung Probabilitas Posterior ($P(C|X)$)

- a. Probabilitas posterior dihitung berdasarkan perkalian probabilitas likelihood dan prior, kemudian dibagi dengan probabilitas evidence.
- b. Probabilitas evidence dihitung berdasarkan jumlah seluruh data dalam dataset yang memiliki karakteristik serupa dengan data yang diuji.

- c. Kelas dengan nilai probabilitas posterior tertinggi akan dipilih sebagai hasil klasifikasi.

4. Menentukan Akurasi Model

- a. Setelah mendapatkan hasil klasifikasi, langkah selanjutnya adalah mengukur akurasi metode *Naïve Bayes* untuk menilai seberapa baik model dapat memprediksi kelas data.
- b. Akurasi dihitung menggunakan rumus:

$$Akurasi = \frac{\text{Jumlah Data yang Diklasifikasikan Benar}}{\text{Total Data Keseluruhan}} \times 100\%$$

- c. Semakin tinggi nilai akurasi, semakin baik model dalam melakukan klasifikasi data.

2.7.2 Proses Implementasi Algoritma *Naïve Bayes*

Dalam implementasi metode *Naïve Bayes*, terdapat beberapa tahapan utama yang harus dilakukan:

1. Membaca Data Training
 - a. Data training adalah kumpulan data yang digunakan untuk melatih model agar dapat melakukan klasifikasi dengan baik.
2. Menghitung Jumlah Data dalam Setiap Kelas
 - a. Model menghitung jumlah data yang masuk dalam masing-masing kategori atau kelas.
3. Menghitung Probabilitas untuk Setiap Kelas
 - a. Setiap kelas dihitung probabilitasnya berdasarkan data yang tersedia dalam dataset.
4. Menghitung Probabilitas Likelihood untuk Setiap Atribut

- a. Model menentukan probabilitas dari masing-masing atribut dalam kelas tertentu.

5. Menghitung Probabilitas Posterior dan Melakukan Klasifikasi

- a. Berdasarkan hasil perhitungan probabilitas likelihood dan prior, model menentukan kelas dengan probabilitas tertinggi sebagai hasil klasifikasi.

2.7.3 Kelebihan Metode *Naïve Bayes*

Metode *Naïve Bayes* memiliki beberapa keunggulan yang menjadikannya metode klasifikasi yang populer, terutama dalam bidang medis dan deteksi penyakit. Berikut beberapa keunggulannya:

1. Mudah Dipahami dan Diterapkan

- a. *Naïve Bayes* memiliki konsep dasar yang sederhana sehingga mudah dipahami dan diimplementasikan, bahkan bagi pengguna yang baru mempelajari *machine learning*.

2. Cepat dalam Perhitungan

- a. Metode ini memiliki waktu komputasi yang rendah, karena hanya melibatkan perhitungan probabilitas sederhana, tanpa perlu melakukan pemodelan kompleks seperti metode lain.

3. Memerlukan Data Latih yang Relatif Sedikit

- a. Berbeda dengan algoritma *machine learning* lainnya, *Naïve Bayes* tetap dapat memberikan hasil klasifikasi yang baik meskipun jumlah data training tidak terlalu banyak.

4. Efisien dalam Ruang Penyimpanan

- a. Algoritma ini tidak memerlukan struktur data yang kompleks, sehingga dapat berjalan dengan kebutuhan ruang memori yang minimal.

5. Cocok untuk Klasifikasi Data dengan Banyak Atribut

- a. Karena metode ini mengasumsikan independensi antar atribut, maka dapat bekerja dengan baik pada dataset yang memiliki banyak fitur atau atribut.

6. Dapat Digunakan pada Berbagai Jenis Data

- a. *Naïve Bayes* dapat digunakan baik untuk data numerik maupun data kategorikal, sehingga fleksibel dalam penerapannya.

2.8 RapidMiner

RapidMiner adalah perangkat lunak open-source yang digunakan untuk analisis data mining dan dirancang untuk mempermudah pengolahan data secara otomatis dengan berbagai metode yang tersedia (Huda, 2019). Dikembangkan sebagai pengembangan dari aplikasi Yet Another Learning Environment (YALE) oleh Ralf Klinberg, Ingo Mierswa, dan Simon Fischer pada tahun 2001, RapidMiner berbasis bahasa pemrograman Java, sehingga dapat digunakan di berbagai sistem operasi. Dalam penelitian ini, RapidMiner digunakan untuk mengolah data rekam medis pasien guna mendukung sistem deteksi dini *Diabetes Melitus* (DM) menggunakan metode *Naïve Bayes*. Perangkat lunak ini memiliki berbagai operator yang mendukung tahap *preprocessing* data, klasifikasi, visualisasi hasil, serta evaluasi model, sehingga memungkinkan analisis data yang lebih akurat dan efisien dalam mengidentifikasi risiko *Diabetes Melitus* (DM) berdasarkan rekam medis pasien.

2.9 Teknik Pemilihan Sampel Dalam Algoritma *Naïve Bayes*

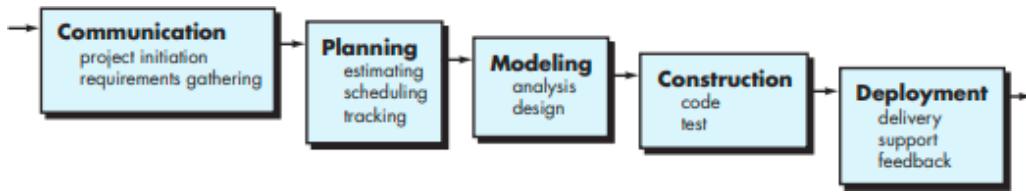
Dalam analisis data mining, pemilihan sampel merupakan langkah penting dalam membangun model klasifikasi yang akurat (Nurussakinah, 2024), termasuk dalam implementasi algoritma *Naïve Bayes* untuk sistem deteksi dini *Diabetes Melitus* (DM) berbasis rekam medis. Teknik sampling yang tersedia dalam perangkat lunak RapidMiner mencakup *linear sampling*, *shuffle sampling*, dan *stratified sampling*.

Linear sampling memilih data uji (*testing*) berdasarkan urutan dalam dataset, di mana sampel uji dipilih dari baris akhir dengan jumlah yang disesuaikan dengan proporsi yang telah ditentukan. Sementara itu, *shuffle sampling* memilih data uji secara acak tanpa memperhatikan distribusi kelas dalam dataset, namun tetap mempertahankan jumlah sampel sesuai pembagian yang digunakan.

Teknik *stratified sampling*, yang merupakan metode paling umum dalam klasifikasi berbasis rekam medis, memilih sampel uji secara acak tetapi tetap mempertimbangkan keseimbangan proporsi antar kelas dalam dataset. Sebagai contoh, jika rasio data pelatihan dan data uji ditetapkan 80:20, maka teknik ini akan memastikan bahwa 20% dari setiap kelas dalam dataset tetap terwakili dalam data uji. Penggunaan teknik stratifikasi ini penting dalam sistem deteksi dini *Diabetes Melitus* (DM), karena memastikan bahwa model tetap seimbang dalam mengenali kasus positif DM dan negatif DM, sehingga meningkatkan keakuratan hasil klasifikasi.

2.10 Metode *Waterfall*

Metode *Waterfall* adalah pendekatan sistematis dalam pengembangan perangkat lunak yang mengikuti tahapan secara berurutan dan linier, di mana setiap tahap harus diselesaikan sebelum melanjutkan ke tahap berikutnya (Juventauricula et al., 2024). Metode ini dikenal sebagai *classic life cycle*, karena setiap proses dilakukan secara bertahap dari perencanaan awal hingga pemeliharaan sistem (Prabowo, 2020). Dalam penelitian ini, metode *Waterfall* digunakan untuk merancang dan mengimplementasikan sistem deteksi dini *Diabetes Melitus* (DM) berbasis rekam medis menggunakan algoritma *Naïve Bayes*, sehingga pengembangan sistem berjalan lebih terstruktur dan efisien (Wahid & Abdul., 2020).



Gambar 2. 1 Metode *Waterfall* (Pressman et al., 2015)

Berikut adalah tahapan utama dalam metode *Waterfall* yang diterapkan dalam penelitian ini:

a. *Communication* (Komunikasi)

Tahapan ini bertujuan untuk mendefinisikan kebutuhan sistem dengan melakukan analisis terhadap permasalahan yang ada. Dalam konteks penelitian ini, komunikasi dilakukan dengan pihak terkait, seperti tenaga medis dan pengelola rekam medis, untuk mengumpulkan informasi tentang variabel yang berpengaruh dalam deteksi dini *Diabetes Melitus* (DM).

b. *Planning* (Perencanaan)

Pada tahap ini, dilakukan penyusunan rencana pengembangan sistem, termasuk identifikasi tugas-tugas yang akan dikerjakan, penjadwalan implementasi, serta alokasi sumber daya yang diperlukan untuk membangun sistem deteksi dini yang berbasis data rekam medis.

c. *Modeling* (Perancangan)

Tahap ini melibatkan perancangan arsitektur sistem, termasuk pembuatan struktur data, desain antarmuka pengguna, dan pemodelan algoritma *Naïve Bayes* yang akan digunakan untuk klasifikasi data rekam medis pasien. Model ini bertujuan untuk memberikan gambaran yang jelas tentang bagaimana sistem akan bekerja dalam memprediksi risiko *Diabetes Melitus* (DM).

d. *Construction* (Konstruksi dan Implementasi)

Tahapan ini merupakan proses pengkodean atau pengembangan perangkat lunak berdasarkan desain yang telah dibuat sebelumnya. Selain itu, dilakukan pengujian terhadap sistem untuk memastikan bahwa algoritma *Naïve Bayes* dapat mengolah

data rekam medis dengan benar serta menghasilkan prediksi yang akurat terhadap risiko *Diabetes Melitus* (DM).

e. *Deployment* (Implementasi dan Pemeliharaan)

Tahap terakhir dalam metode *Waterfall* adalah implementasi sistem kepada pengguna, yaitu tenaga medis atau rumah sakit yang akan menggunakan sistem deteksi dini *Diabetes Melitus* (DM) ini. Setelah sistem diterapkan, dilakukan pemeliharaan berkala, evaluasi performa, serta pengembangan lebih lanjut berdasarkan umpan balik dari pengguna, guna memastikan sistem tetap berfungsi dengan baik dan dapat memberikan hasil prediksi yang optimal.

2.11 *Confusion Matrix*

Confusion Matrix merupakan salah satu metode evaluasi yang digunakan dalam analisis data mining, khususnya dalam teknik klasifikasi, termasuk dalam penerapan algoritma *Naïve Bayes* untuk sistem deteksi dini *Diabetes Melitus* (DM) berbasis rekam medis. *Confusion Matrix* disajikan dalam bentuk tabel yang menggambarkan jumlah data yang diprediksi dengan benar serta jumlah data yang diklasifikasikan secara salah oleh model (Fikri et al., 2020). Dengan menggunakan *Confusion Matrix*, kinerja sistem klasifikasi dapat diukur berdasarkan berbagai metrik evaluasi, seperti *accuracy*, *precision*, dan *recall*.

Tabel berikut menunjukkan format dasar *Confusion Matrix* untuk sistem klasifikasi dua kelas, yaitu positif *Diabetes Melitus* (DM) dan negatif DM:

Tabel 2. 2 *Confusion Matrix*

		Kelas hasil prediksi	
		Kelas berisiko	Kelas tidak berisiko
Kelas	Kelas berisiko	TP	FN
	Kelas tidak berisiko	FP	TN

- a. *True Positive* (TP) sesuai dengan jumlah data bernilai positif yang telah diprediksi benar sebagai nilai positif. Contohnya, jumlah label "BERESIKO" yang terdeteksi dengan benar.

- b. *False Positive* (FP) berhubungan dengan jumlah data negatif yang diprediksi sebagai positif. Misalnya, jumlah label "TIDAK BERESIKO" yang diprediksi salah, oleh karena itu dinyatakan sebagai label "BERESIKO".
- c. *True Negative* (TN) adalah jumlah data negatif yang diprediksi negatif. Sebagai contoh, jumlah label "TIDAK BERESIKO" yang diprediksi benar.
- d. *False Negative* (FN) adalah jumlah data yang memiliki nilai positif tetapi diprediksi sebagai data negatif, contohnya, jumlah label "BERESIKO" yang diprediksi salah sehingga dinyatakan "TIDAK BERESIKO".

Pada prakteknya *Confusion Matrix* digunakan untuk menghitung nilai *accuracy*, *precision*, dan *recall* (Romadloni et al., 2022).

- a. Nilai *Accuracy*, menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar. Nilai ini diperoleh dari perbandingan antara jumlah prediksi yang benar (baik positif maupun negatif) terhadap keseluruhan jumlah data. Semakin tinggi nilai akurasi, semakin baik kinerja model dalam melakukan klasifikasi secara umum.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

- b. *Precision*, menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model. *Precision* merupakan rasio antara jumlah data positif yang diprediksi dengan benar terhadap seluruh data yang diprediksi sebagai positif. Nilai *precision* yang tinggi menunjukkan bahwa sistem memiliki kecenderungan rendah dalam menghasilkan kesalahan jenis *false positive*, yaitu memprediksi pasien berisiko padahal sebenarnya tidak.

$$\text{Precision} = (\text{TP}) / (\text{TP} + \text{FP})$$

- c. *Recall*, menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. *Recall* dihitung dengan membandingkan jumlah *true positive* terhadap jumlah total data aktual positif (*true positive* + *false negative*). Nilai *recall* yang tinggi sangat penting karena menunjukkan bahwa sistem mampu mendeteksi sebagian besar pasien berisiko tanpa banyak yang terlewat.

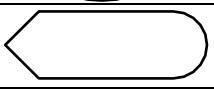
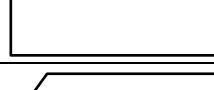
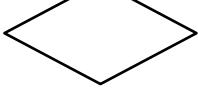
$$Recall = TP / (TP + FN)$$

2.12 Flowchart

Flowchart adalah representasi grafis dari alur kerja suatu sistem atau program yang disusun secara sistematis agar lebih mudah dipahami (Fauzi, 2020). *Flowchart* berfungsi sebagai panduan visual untuk menggambarkan proses yang terjadi dalam sistem, seperti langkah-langkah dalam sistem deteksi dini *Diabetes Melitus* (DM) berbasis rekam medis. Jika dalam algoritma alur proses dituliskan dalam bentuk baris kode pemrograman, maka dalam *flowchart* alur tersebut divisualisasikan menggunakan simbol-simbol tertentu yang merepresentasikan berbagai tahapan dalam sistem (Firdaus, 2021).

Berikut adalah beberapa simbol yang umum digunakan dalam *flowchart*:

Tabel 2. 3 Simbol *Flowchart*

Nama simbol	Simbol	Fungsi
<i>Keyboard</i> (manual input)		Input yang menggunakan keyboard
<i>Punched card</i>		Input / output yang menggunakan kartu
Harddisk		Input / output yang menggunakan harddisk
<i>Display</i>		Output yang ditampilkan di layar
Dokumen		Dokumen input dan output baik untuk proses manual ataupun otomatis
Terminator		Memulai dan mengakhiri program
Proses		Menunjukkan operasi yang sedang diproses
<i>Read / write</i>		Menunjukkan input ataupun output
<i>Decision</i>		Menunjukkan proses pemeriksaan atau evaluasi terhadap suatu kondisi

Nama simbol	Simbol	Fungsi
<i>Preparation</i>		Menunjukkan pemberian atau pendeklarasian variabel atau nilai awal
Sub program		Menunjukkan sub program dari program utama yang akan dijalankan
<i>Connector</i>		Menunjukkan tanda penghubung flowchart yang masih satu halaman
<i>Off page connector</i>		Menunjukkan tanda penghubung flowchart pada halaman yang berbeda

2.13 Data Flow Diagram

Data Flow Diagram (DFD) adalah representasi aliran data dalam suatu sistem. DFD menggambarkan pergerakan data dari input hingga menghasilkan output. Diagram ini divisualisasikan menggunakan simbol-simbol tertentu yang dapat dilihat pada Tabel 2.4. DFD berfungsi sebagai alat pemodelan yang digunakan dalam perancangan sistem yang berfokus pada aliran data. Selain itu, DFD juga berperan sebagai sarana komunikasi antara profesional sistem dengan pengguna atau pengembang program (Ali, 2019).

Tabel 2. 4 Simbol *Data Flow Diagram*

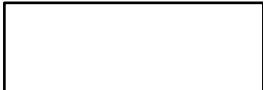
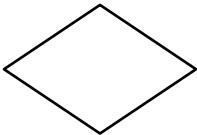
Nama simbol	Simbol	Keterangan
Proses		Menggambarkan proses yang akan diimplementasikan pada sistem.
<i>Data store</i>		Menggambarkan penyimpanan atau database yang ada pada sistem.
<i>Source / sink</i>		Menggambarkan masukan, hasil keluaran, entitas luar, atau orang yang menggunakan sistem.
<i>Data flow</i>		Menggambarkan arah aliran data yang terjadi pada sistem.

2.14 Entity Relationship Diagram

Entity Relationship Diagram (ERD) adalah diagram yang merepresentasikan hubungan dalam basis data berdasarkan fakta dari entitas di dunia nyata. Fakta-fakta tersebut dipetakan ke dalam model ERD, yang mencakup kumpulan entitas beserta atribut-atributnya serta relasi antar entitas. Tabel 2.4 menunjukkan simbol-simbol yang digunakan dalam diagram ERD (Ali, 2019).

Tabel 2. 5 Simbol Entity Relationship Diagram

Nama simbol	Simbol	Keterangan
-------------	--------	------------

Entitas		Menggambarkan objek yang dapat diidentifikasi dari dunia nyata
Relasi		Menggambarkan hubungan diantara entitas
Atribut		Menggambarkan atau mendeskripsikan karakter dari entitas
Garis		Sebagai penghubung antara relasi dengan entitas dan entitas dengan atribut

Sumber: (Solikin, Sobri, and Saputra 2018)

2.15 Pemrograman Web

Website merupakan sekumpulan halaman digital yang berisi informasi dalam bentuk teks, gambar, animasi, audio, atau video, serta kombinasi dari berbagai elemen tersebut yang dapat diakses melalui internet. Pemrograman web adalah proses pembuatan aplikasi berbasis web menggunakan bahasa skrip tertentu agar dapat diakses melalui peramban (*web browser*). Beberapa bahasa pemrograman yang umum digunakan dalam pengembangan website meliputi HTML, PHP, CSS, dan JavaScript (Sari & Abdilah, 2019).

2.16 *Hypertext Preprocessor*

PHP (*Hypertext Preprocessor*) adalah bahasa pemrograman yang digunakan untuk membangun website dinamis, interaktif, dan berfungsi sebagai aplikasi berbasis web yang lengkap. PHP memiliki ekstensi file .php dan dapat digunakan untuk berbagai keperluan, seperti mengakses basis data, membuat

gambar, serta membaca dan menulis file. Kemampuannya dalam menangani berbagai proses dinamis menjadikannya salah satu bahasa pemrograman yang banyak digunakan dalam pengembangan aplikasi web (Sari & Abdilah, 2019)

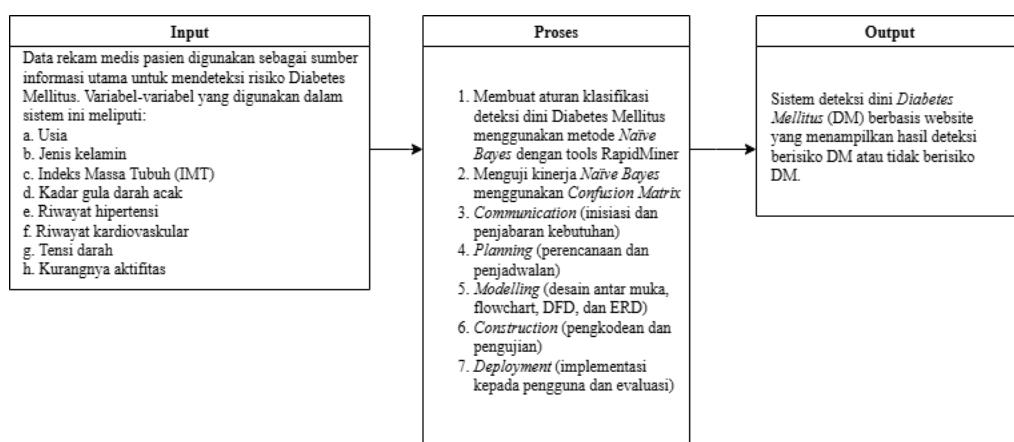
2.17 Blackbox Testing

Blackbox Testing adalah salah satu metode pengujian sistem yang berfokus pada pemeriksaan output berdasarkan input yang diberikan, tanpa memperhatikan detail internal atau kode program yang digunakan. Metode ini bertujuan untuk mengevaluasi apakah sistem telah berfungsi sesuai dengan kebutuhan dan spesifikasi yang telah ditentukan, tanpa perlu menganalisis struktur atau logika pemrograman yang ada (Ningrum et al., 2019)

2.18 Sistem Deteksi Dini

Sistem deteksi dini merupakan sistem yang dirancang untuk mengidentifikasi kasus suatu penyakit sejak tahap awal di masyarakat. Tujuan utama dari sistem ini adalah mencegah penyebaran penyakit, terutama jika penyakit tersebut bersifat menular (Putra et al., 2019). Deteksi yang lebih cepat memungkinkan pasien memperoleh pengobatan lebih awal, sehingga dapat mencegah perkembangan penyakit ke kondisi yang lebih serius (Panna et al., 2021).

2.19 Kerangka Konsep



Gambar 2. 2 Kerangka Konsep

Kerangka konsep dalam sistem deteksi dini penyakit *Diabetes Melitus* (DM) menggunakan metode *Naïve Bayes* berbasis data rekam medis terdiri dari tiga tahap utama, yaitu input, proses, dan output. Pada tahap input, dilakukan pengumpulan

data rekam medis pasien yang mencakup usia, jenis kelamin, indeks massa tubuh (IMT), kadar gula darah acak, tensi darah, riwayat hipertensi, riwayat kardiovaskular serta kurangnya aktivitas yang diperoleh dari rekam medis Rumah Sakit Citra Husada Jember. Pembuatan aturan klasifikasi menggunakan algoritma *Naïve Bayes* dengan RapidMiner. Selanjutnya pengujian performa hasil klasifikasi algoritma *Naïve Bayes* menggunakan *Confusion Matrix* dengan RapidMiner. *Communication* dilakukan untuk pendeskripsian kebutuhan sistem, setelah itu *Planning* pengembangan sistem, perancangan antarmuka pengguna, *flowchart*, *Data Flow Diagram* (DFD), dan *Entity-Relationshipship Diagram* (ERD), dan terakhir terdiri dari pengkodean dan pengujian perangkat lunak. Pada tahap output, sistem menghasilkan deteksi dini *Diabetes Mellitus* (DM), dengan dua kategori yaitu Berisiko DM jika pasien memiliki kemungkinan besar terkena DM, dan Tidak Berisiko DM jika pasien tidak menunjukkan risiko signifikan terhadap penyakit tersebut.

BAB 3. METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif yang menggunakan metode klasifikasi dalam data mining untuk membangun sistem deteksi dini *Diabetes Mellitus* (DM). Algoritma yang digunakan dalam penelitian ini adalah *Naïve Bayes* sebagai metode utama untuk klasifikasi data rekam medis pasien. Selain itu, penelitian ini juga menggunakan metode *Waterfall* dalam pengembangan sistem berbasis web.

3.2 Waktu dan Tempat Penelitian

Penelitian ini dilakukan dalam rentang waktu Juni 2025 hingga Agustus 2025. Penelitian ini dilakukan di RS Citra Husada, Kabupaten Jember. Data yang digunakan dalam penelitian ini diperoleh dari rekam medis pasien.

3.3 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini ditentukan berdasarkan faktor risiko yang berkontribusi dalam diagnosis *Diabetes Mellitus* berdasarkan literatur yang relevan. Berikut adalah variabel penelitian yang digunakan:

a. Variabel Input (Faktor Risiko DM):

1. Usia (Dewasa awal/Dewasa akhir/Lansia awal/Lansia akhir/Manula)
2. Jenis Kelamin (Laki-laki/Perempuan)
3. Indeks Massa Tubuh (Kurus/Normal/Gemuk)
4. Kadar Gula Darah Acak (Berisiko DM/Tidak Berisiko DM)
5. Tensi Darah (Normal/Pra-hipertensi/Hipertensi tingkat I/Hipertensi tingkat II)
6. Riwayat Hipertensi (Ya/Tidak)
7. Riwayat Kardiovaskular (Ya/Tidak)
8. Kurangnya Aktivitas (Cukup/Kurang)

b. Variabel Output (Hasil Klasifikasi):

1. Berisiko DM
2. Tidak Berisiko DM

3.4 Objek Penelitian

Objek penelitian ini adalah rekam medis pasien di RS Citra Husada, Kabupaten Jember. Data yang digunakan mencakup rawat inap dengan diagnosis *Diabetes Mellitus* berdasarkan ICD-10 E10-E14.

3.5 Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah observasi data sekunder dengan melakukan ekstraksi informasi dari rekam medis pasien. Data yang dikumpulkan mencakup variabel yang telah ditentukan, seperti usia, jenis kelamin, kadar gula darah acak, IMT, tensi darah, riwayat hipertensi, riwayat kardiovaskular, dan kurangnya aktivitas.

3.6 Instrumen Penelitian

Instrumen penelitian yang digunakan berupa lembar observasi dalam format Excel yang mencatat variabel-variabel yang digunakan dalam sistem klasifikasi. Data dikodekan dalam bentuk kategori tertentu agar dapat digunakan dalam model klasifikasi *Naïve Bayes*.

3.7 Definisi Operasional

Berikut adalah beberapa istilah penting dalam penelitian ini:

Tabel 3. 1 Definisi Operasional

No	Variabel	Definisi Operasional	Cara Pengumpulan Data	Kriteria Pengukuran
1.	Usia	Informasi tentang usia pasien yang dapat digunakan dalam analisis risiko <i>Diabetes Melitus</i> . Usia dikategorikan menjadi lima diantaranya yaitu masa dewasa awal (26-35 tahun), masa dewasa akhir (36-45 tahun), masa lansia awal (46-55 tahun), masa lansia akhir (56-65 tahun), masa manula (> 65 tahun) (Katmawanti et al., 2022)	Observasi data rekam medis pasien	1. Masa Dewasa Awal (26-35 Tahun) 2. Masa Dewasa Akhir (36-45 Tahun) 3. Masa Lansia Awal (46-55 Tahun) 4. Masa Lansia Akhir (56-65 Tahun) 5. Masa Manula (> 65 Tahun)
2.	Indeks Massa Tubuh (IMT)	Parameter yang digunakan untuk mengklasifikasikan berat badan pasien	Observasi data rekam medis pasien	1. Kurus (<18,4) 2. Normal (18,5-25,0)

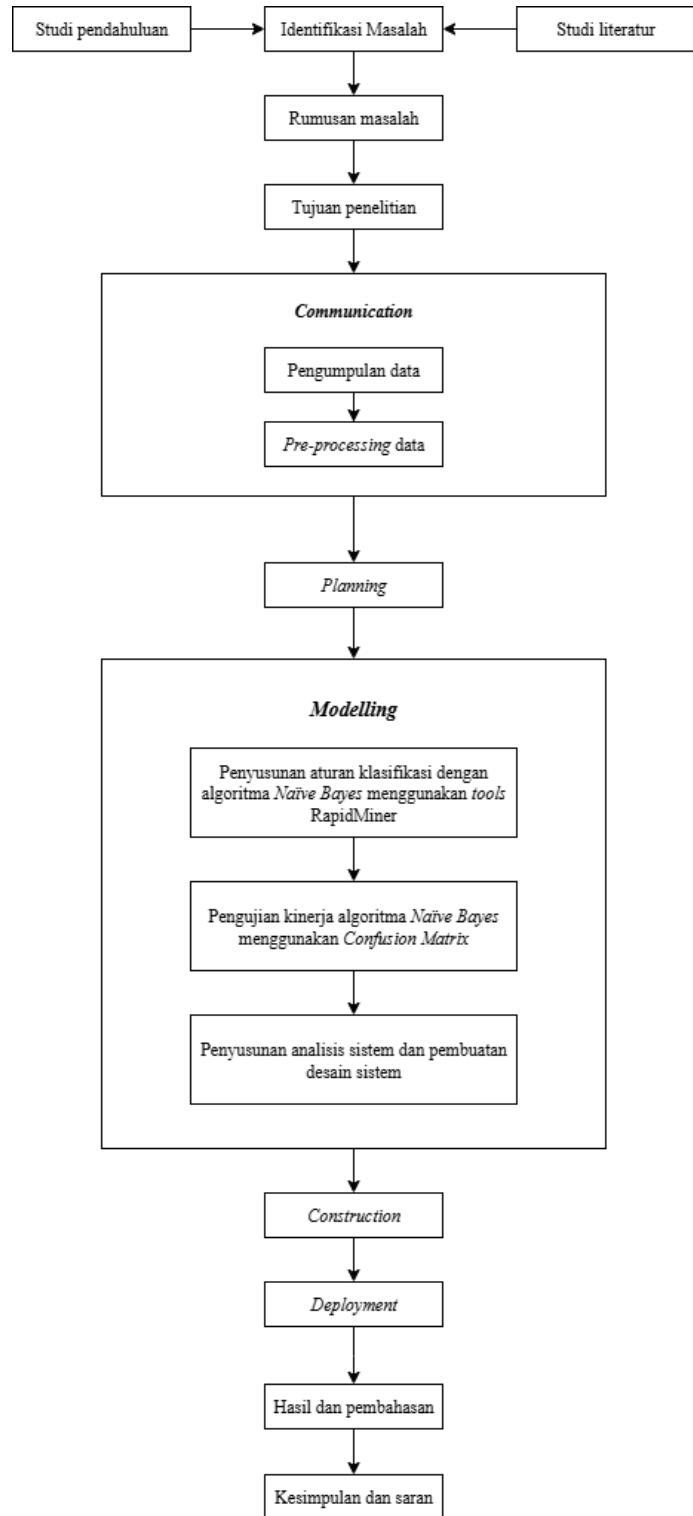
No	Variabel	Definisi Operasional	Cara Pengumpulan Data	Kriteria Pengukuran
		guna mengetahui kaitannya dengan risiko <i>diabetes mellitus</i> .		3. Gemuk ($>25,1$)
		Kategori Indeks Massa Tubuh (IMT) diantaranya kurus ($<18,4$), normal ($18,5-25,0$), gemuk ($>25,1$) (P2PTM Kemenkes RI)		
3.	Jenis Kelamin	Kategori biologis yang membedakan individu berdasarkan karakteristik fisik pasien dan genetik, biasanya diklasifikasikan sebagai laki-laki atau perempuan.	Observasi data rekam medis	1. Perempuan 2. Laki-laki
4.	Kadar Gula Darah Acak	Kadar glukosa dalam darah yang diukur tanpa memperhatikan waktu terakhir kali pasien seseorang makan. Diukur dalam mg/dL. Tes ini mengukur gula darah secara acak pada saat tes dilakukan dan tidak memerlukan puasa terlebih dahulu.	Observasi data rekam medis	1. Normal (<132 mg/dL) 2. Berisiko (>132 mg/dL)

No	Variabel	Definisi Operasional	Cara Pengumpulan Data	Kriteria Pengukuran
		Berdasarkan standar klinis rumah sakit, jika hasil gula darah acak mencapai 132 mg/dL atau lebih, pasien berisiko mengalami diabetes melitus.		
5.	Tensi Darah	<p>Ukuran tekanan darah dalam arteri, dinyatakan dalam mmHg, dengan dua angka: tekanan sistolik (tekanan saat jantung berkontraksi) dan diastolik (tekanan saat jantung beristirahat).</p> <p>Sesuai dengan kesepakatan JNC7 yaitu tujuh panduan dalam klasifikasi dan jenis terapi hipertensi versi internasional maka pengelompokan tekanan darah yaitu normal (<120/80), pra-hipertensi (120/85-139/89), hipertensi tingkat I (140/90-159/99), hipertensi tingkat II ($\geq 160/100$) (Nurrahmani, 2015).</p>	<p>Observasi data rekam medis</p>	<ol style="list-style-type: none"> 1. Normal (<120/80) 2. Pra-Hipertensi (120/85-139/89) 3. Hipertensi Tingkat I (140/90-159/99) 4. Hipertensi Tingkat II ($\geq 160/100$)

No	Variabel	Definisi Operasional	Cara Pengumpulan Data	Kriteria Pengukuran
6.	Riwayat Hipertensi	Rekam medis yang menunjukkan bahwa individu pernah didiagnosis dengan pasien hipertensi, yaitu kondisi di mana tekanan darah secara konsisten berada di atas batas normal.	Observasi data rekam medis	1. Ya (Memiliki Riwayat Hipertensi) 2. Tidak (Tidak Memiliki Riwayat Hipertensi)
7.	Riwayat Kardiovaskular	Rekam medis yang menunjukkan adanya penyakit atau kondisi yang mempengaruhi pasien jantung dan pembuluh darah, seperti serangan jantung, stroke, atau penyakit jantung koroner.	Observasi data rekam medis	1. Ya (Memiliki Riwayat Kardiovaskular) 2. Tidak (Tidak Memiliki Riwayat Kardiovaskular)
8.	Kurangnya Aktivitas	Kondisi dimana individu tidak memenuhi rekomendasi aktivitas fisik yang pasien disarankan, biasanya kurang dari 150 menit aktivitas fisik moderat per minggu.	Observasi data rekam medis	1. Kurang 2. Cukup

3.8 Tahapan Penelitian

Tahapan dalam penelitian ini mengikuti langkah-langkah berikut:



Gambar 4. 1 Tahapan penelitian

Penjelasan dari tahapan penelitian adalah sebagai berikut:

a. Studi pendahuluan

Studi pendahuluan merupakan langkah awal dalam menyelidiki permasalahan yang berkaitan dengan deteksi dini *diabetes mellitus*. Sebagai bagian dari kegiatan ini, data dan informasi yang relevan dari sumber-sumber resmi seperti Organisasi Kesehatan Dunia (WHO), Kementerian Kesehatan Indonesia dan laporan kesehatan Rumah Sakit Citra Husada diteliti untuk memahami pentingnya sistem deteksi dini *diabetes mellitus* berdasarkan data dari rekam medis.

Selain mengumpulkan informasi dari sumber-sumber yang dapat dipercaya, studi pendahuluan ini juga bertujuan untuk memahami pola dan tren *diabetes mellitus* pada berbagai kelompok populasi. Analisis awal ini dapat memberikan wawasan tentang faktor risiko yang sering terjadi sehingga sistem yang dikembangkan dapat lebih relevan dan tepat sasaran dalam melakukan deteksi dini.

b. Studi literatur

Studi literatur dilakukan untuk mengumpulkan teori, konsep, dan metode yang relevan dengan penelitian, terutama yang berkaitan dengan algoritma *Naïve Bayes*, sistem deteksi penyakit, dan penggunaan data rekam medis dalam klasifikasi penyakit. Sumber-sumber tersebut diambil dari jurnal ilmiah, buku, laporan akademik dan hasil penelitian sebelumnya.

Studi literatur juga memungkinkan para peneliti untuk membandingkan berbagai metode klasifikasi yang digunakan dalam penelitian sebelumnya. Hal ini memungkinkan untuk memilih metode yang paling sesuai.

c. Identifikasi masalah

Tujuan dari tahap ini adalah untuk merumuskan permasalahan utama penelitian, yaitu kurangnya sistem yang mampu membantu staf medis mendeteksi risiko *diabetes mellitus* dengan cara yang cepat dan terintegrasi. Masalah ini dirinci dalam rumusan masalah, tujuan penelitian, dan manfaat penelitian.

Pada tahap ini, peneliti juga dapat melakukan kuesioner atau wawancara dengan staf medis dan pasien untuk mendapatkan informasi langsung mengenai tantangan yang terkait dengan deteksi dini *diabetes mellitus*. Informasi ini berguna untuk merancang sistem yang benar-benar memenuhi kebutuhan pengguna di dunia nyata.

d. Rumusan masalah

Rumusan masalah disusun dalam bentuk pertanyaan yang harus dijawab dalam penelitian, "Bagaimana merancang dan mengimplementasikan sebuah sistem deteksi dini penyakit *diabetes* menggunakan algoritma *Naïve Bayes* berbasis *website*?". Permasalahan harus dirumuskan dengan jelas dan spesifik agar penelitian dapat fokus pada satu titik tertentu.

e. Tujuan penelitian

Tujuan penelitian ini adalah untuk menghasilkan sistem deteksi dini *diabetes mellitus* yang memanfaatkan data dari rekam medis dan dapat mengklasifikasikan risiko *diabetes mellitus* secara efisien dan akurat dengan menggunakan algoritma *Naïve Bayes*.

Selain mengembangkan sistem deteksi dini yang akurat, tujuan penelitian juga mencakup pengoptimalan efisiensi kerja staf medis. Dengan sistem yang berbasis data dari rekam medis, diagnosis dini dapat dilakukan dengan lebih cepat sehingga pasien yang berisiko terkena *diabetes mellitus* dapat ditangani dengan lebih proaktif.

f. Pengumpulan data

Data yang dikumpulkan berasal dari rekam medis pasien, yang meliputi variabel seperti usia, jenis kelamin, indeks massa tubuh, kadar glukosa darah acak, tekanan darah, riwayat hipertensi, riwayat kardiovaskular, dan kurangnya aktivitas. Data ini berfungsi sebagai dasar untuk melatih dan menguji model *Naïve Bayes*.

Pada tahap pengumpulan data, penting untuk memastikan bahwa data yang digunakan telah melalui proses validasi untuk menghindari bias pada model. Data yang dikumpulkan harus mencerminkan berbagai kondisi *diabetes mellitus* sehingga sistem dapat memiliki tingkat generalisasi yang baik.

g. *Pre-processing* data

Tahap ini melibatkan pembersihan data dari nilai nol atau tidak valid, mengubah data numerik menjadi bentuk kategorikal dan memilih atribut yang relevan. Hal ini penting untuk memastikan bahwa model Naïve Bayes memiliki data yang bersih dan konsisten.

Selain pembersihan data dan konversi ke bentuk kategorikal, tahap *pre-processing* juga dapat mencakup teknik normalisasi dan pengurangan dimensi. Hal ini akan meningkatkan efisiensi model klasifikasi dalam mendeteksi pola diabetes berdasarkan berbagai variabel yang tersedia.

h. Penyusunan Aturan Klasifikasi

Tujuan dari tahap ini adalah membuat model untuk mengklasifikasikan risiko *Diabetes Melitus* (DM) dengan menggunakan algoritma *Naïve Bayes*. Data riwayat pasien diproses menggunakan RapidMiner untuk membuat aturan klasifikasi berbasis probabilitas.

Kinerja model kemudian dievaluasi dengan menggunakan *Confusion Matrix* untuk mengukur *accuracy*, *precision* dan *recall*. Evaluasi ini penting untuk memastikan bahwa sistem dapat mengidentifikasi pasien berisiko dengan benar, terutama untuk menghindari kesalahan negatif palsu dalam deteksi dini.

i. Pembuatan Sistem

Sistem deteksi dini kemudian dikembangkan dengan menggunakan pendekatan waterfall yang terdiri dari lima tahap. Tahap pertama adalah communication, yaitu proses mengumpulkan kebutuhan sistem melalui studi pendahuluan dan diskusi dengan pihak terkait. Selanjutnya adalah planning atau perencanaan pengembangan sistem, mencakup penjadwalan dan perancangan alur kerja. Tahap berikutnya adalah modelling, yang mencakup perancangan struktur sistem seperti *flowchart*, *Data Flow Diagram* (DFD), dan *Entity Relationship Diagram* (ERD). Setelah itu, masuk ke tahap construction, yaitu proses pengkodean program dan implementasi sistem berbasis web. Terakhir, dilakukan deployment, yakni pengujian sistem melalui simulasi data pengguna untuk memastikan fungsionalitas dan keakuratan klasifikasi risiko

Diabetes Mellitus. Hasil dari tahap ini adalah sistem deteksi dini DM berbasis web yang terintegrasi dengan model klasifikasi *Naïve Bayes*.

j. Analisis

Tahapan analisis dilakukan untuk mengevaluasi performa model klasifikasi. Metode evaluasi yang digunakan adalah *Confusion Matrix*, dengan pengukuran *accuracy*, *precision*, dan *recall* untuk menilai kinerja prediksi sistem.

Evaluasi performa model tidak hanya melihat akurasi, tetapi juga menilai bagaimana sistem menangani data yang tidak seimbang. Jika terdapat lebih banyak data negatif dibandingkan positif, maka strategi seperti oversampling atau penyesuaian bobot mungkin diperlukan agar prediksi tetap akurat.

k. Pembuatan sistem

Sistem dikembangkan menggunakan model Waterfall. Tahapan meliputi analisis kebutuhan, perancangan antarmuka, pengkodean logika *Naïve Bayes*, dan integrasi data. Sistem ini dirancang untuk mempermudah pengguna dalam mendeteksi risiko *diabetes mellitus* melalui input data rekam medis secara langsung.

Dalam pengembangan sistem, aspek user experience juga harus diperhatikan. Antarmuka yang intuitif dan mudah digunakan oleh tenaga medis dapat meningkatkan adopsi sistem secara luas, sehingga manfaat deteksi dini *diabetes mellitus* dapat benar-benar dirasakan dalam praktik medis sehari-hari.

l. Hasil dan pembahasan

Tahapan ini membahas hasil akhir dari implementasi sistem, meliputi tampilan sistem, hasil klasifikasi, serta analisis performa sistem berdasarkan evaluasi yang telah dilakukan. Hasil tersebut disusun dalam bentuk laporan penelitian.

Bagian hasil dan pembahasan juga mencakup perbandingan dengan sistem deteksi dini lainnya. Peneliti akan menginterpretasikan kelebihan dan kekurangan sistem yang dikembangkan serta memberikan wawasan tentang bagaimana sistem ini bisa diperbaiki di masa depan.

m. Kesimpulan dan saran

Penelitian ditutup dengan kesimpulan dari hasil yang diperoleh, serta saran untuk pengembangan sistem di masa depan, seperti penambahan fitur lain atau pengujian lebih lanjut dengan dataset yang lebih besar.

Kesimpulan memberikan ringkasan jelas terkait pencapaian penelitian serta implikasi praktisnya. Saran yang akan diberikan perlu mempertimbangkan aspek teknis dan non-teknis, seperti kemudahan integrasi dengan sistem kesehatan yang sudah ada dan potensi pengembangan dengan teknologi terbaru seperti machine learning yang lebih canggih.

3.9 Pengujian Akurasi dengan *Confusion Matrix*

Pengujian akurasi dengan *Confusion Matrix* merupakan salah satu metode pengujian yang digunakan dalam analisis data mining, khususnya dalam teknik klasifikasi, termasuk dalam penerapan algoritma *Naïve Bayes* untuk sistem deteksi dini *Diabetes Melitus* (DM) berbasis rekam medis. Evaluasi model dilakukan menggunakan *Confusion Matrix* untuk mengukur kinerja klasifikasi dengan metrik berikut:

- True Positive* (TP): Jumlah kasus DM yang terdeteksi dengan benar.
- True Negative* (TN): Jumlah kasus non-DM yang terdeteksi dengan benar.
- False Positive* (FP): Kasus yang diklasifikasikan sebagai DM padahal tidak.
- False Negative* (FN): Kasus yang diklasifikasikan sebagai non-DM padahal positif.

Rumus akurasi dihitung sebagai beriku:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

3.10 Map Menu Sistem

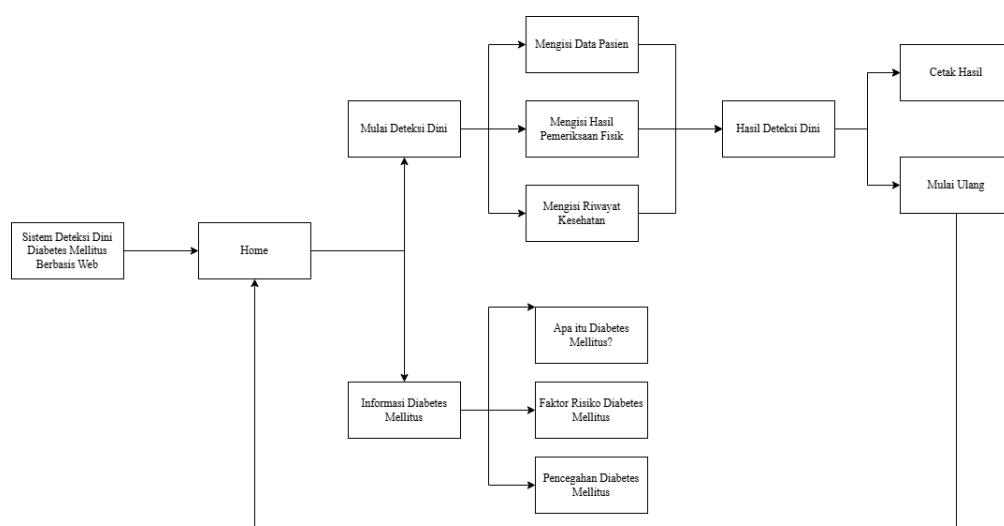
Map menu sistem merupakan gambaran struktur navigasi dari sistem deteksi dini *Diabetes Melitus* berbasis website yang dikembangkan dalam penelitian ini. Struktur menu ini dirancang untuk memudahkan pengguna dalam mengakses informasi mengenai *Diabetes Melitus* (DM), melihat informasi DM, mengisi data pribadi pasien, melakukan deteksi dini, melihat hasil analisis serta mencetak hasil analisis yang diberikan oleh sistem.

Pada sistem ini, halaman utama (Home) menjadi pusat navigasi utama yang menghubungkan ke berbagai menu lain, termasuk informasi Diabetes, registrasi pengguna, dan login pengguna. Menu informasi Diabetes menyediakan sub-menu seperti Definisi Diabetes, Gejala Diabetes, dan Faktor Risiko Diabetes, yang bertujuan untuk memberikan edukasi kepada pengguna sebelum melakukan deteksi dini.

Bagi pengguna yang ingin menggunakan fitur deteksi dini, mereka harus melakukan registrasi terlebih dahulu, kemudian masuk melalui menu Login Pengguna. Setelah masuk, mereka dapat mengakses fitur Deteksi Dini Diabetes, melihat Riwayat Deteksi yang telah dilakukan sebelumnya, atau keluar dari sistem melalui menu Logout.

Setelah proses deteksi dini dilakukan, sistem akan menampilkan Hasil Deteksi Dini *Diabetes Mellitus*, yang akan memberikan indikasi apakah seorang pasien memiliki risiko terkena *Diabetes Mellitus* atau tidak. Hasil ini diperoleh melalui proses analisis berbasis algoritma *Naïve Bayes* yang bekerja dengan mengolah data rekam medis pasien.

Gambar di bawah ini menggambarkan struktur navigasi sistem deteksi dini *Diabetes Mellitus* berbasis website. Pada diagram ini:



Gambar 4. 2 Map menu sistem

1. Home adalah halaman utama yang menghubungkan ke berbagai fitur dalam sistem. Sistem ini bisa diakses oleh siapa saja, mulai dari pasien dan dokter. Pasien dan dokter dapat mengakses home dan langsung menginput data.
2. Dari Home, pengguna dapat mengakses:

- a) Deteksi dini, yang berisi:
 - Data pasien
 - Hasil pemeriksaan fisik
 - Riwayat kesehatan

Hasil deteksi dini akan muncul setelah melakukan tes diatas dan bisa dicetak atau mulai ulang.

- b) Informasi tentang *diabetes mellitus* yang pengguna dapatkan meliputi:
 - Definisi
 - Faktor risiko
 - pencegahan

Dengan struktur ini, sistem dirancang agar lebih user-friendly dan memudahkan pengguna dalam melakukan deteksi dini *Diabetes Melitus*.

3.11 Contoh Penghitungan Klasifikasi Metode *Naïve Bayes*

Tabel 3. 2 ata training

Umur	Jenis Kelamin	TENSI	IMT	GDA	HT	RIW KAI	AKTIFITAS	DIAGNOSA
LANSIA AWAL	Perempuan	HT TINGKAT II	GEMUK	BERISIK	YA	TIDAK	KURANG	DM
LANSIA AWAL	Perempuan	NORMAL	NORMAL	NORMA	TIDAK	TIDAK	CUKUP	DM
MANULA	Laki-laki	HT TINGKAT I	GEMUK	BERISIK	YA	TIDAK	CUKUP	DM
LANSIA AKHIR	Perempuan	HT TINGKAT II	NORMAL	NORMA	TIDAK	TIDAK	CUKUP	DM
LANSIA AKHIR	Perempuan	NORMAL	NORMAL	NORMA	YA	TIDAK	CUKUP	DM
LANSIA AWAL	Perempuan	NORMAL	NORMAL	BERISIK	TIDAK	TIDAK	KURANG	DM
MANULA	Laki-laki	HT TINGKAT I	NORMAL	NORMA	TIDAK	YA	CUKUP	DM
LANSIA AWAL	Perempuan	NORMAL	NORMAL	BERISIK	TIDAK	TIDAK	CUKUP	DM
LANSIA AKHIR	Laki-laki	NORMAL	GEMUK	NORMA	YA	YA	CUKUP	DM
LANSIA AKHIR	Perempuan	PRA-HIPERTENSI	NORMAL	NORMA	YA	YA	CUKUP	DM
MANULA	Laki-laki	PRA-HIPERTENSI	NORMAL	BERISIK	TIDAK	TIDAK	KURANG	DM
LANSIA AKHIR	Perempuan	HT TINGKAT II	GEMUK	BERISIK	YA	TIDAK	CUKUP	DM
LANSIA AKHIR	Laki-laki	NORMAL	NORMAL	NORMA	TIDAK	TIDAK	CUKUP	DM
LANSIA AKHIR	Laki-laki	HT TINGKAT I	NORMAL	BERISIK	TIDAK	TIDAK	CUKUP	DM
MANULA	Perempuan	HT TINGKAT I	NORMAL	NORMA	TIDAK	TIDAK	CUKUP	NON DM
LANSIA AWAL	Perempuan	HT TINGKAT II	GEMUK	NORMA	YA	TIDAK	KURANG	NON DM
LANSIA AKHIR	Perempuan	PRA-HIPERTENSI	GEMUK	NORMA	YA	TIDAK	CUKUP	NON DM
MANULA	Laki-laki	HT TINGKAT I	NORMAL	BERISIK	YA	TIDAK	CUKUP	NON DM
DEWASA AKHIR	Perempuan	HT TINGKAT I	GEMUK	BERISIK	TIDAK	TIDAK	CUKUP	NON DM
LANSIA AWAL	Perempuan	PRA-HIPERTENSI	NORMAL	NORMA	YA	TIDAK	KURANG	NON DM
DEWASA AWAL	Perempuan	HT TINGKAT II	GEMUK	NORMA	YA	TIDAK	CUKUP	NON DM
LANSIA AKHIR	Perempuan	PRA-HIPERTENSI	GEMUK	NORMA	TIDAK	TIDAK	KURANG	NON DM
DEWASA AKHIR	Perempuan	HT TINGKAT I	GEMUK	NORMA	YA	TIDAK	CUKUP	NON DM
MANULA	Perempuan	HT TINGKAT I	NORMAL	NORMA	TIDAK	TIDAK	CUKUP	NON DM
LANSIA AKHIR	Perempuan	HT TINGKAT II	NORMAL	NORMA	TIDAK	TIDAK	CUKUP	NON DM
LANSIA AKHIR	Laki-laki	HT TINGKAT II	NORMAL	NORMA	TIDAK	TIDAK	CUKUP	NON DM

Langkah pertama adalah mengumpulkan data. Dataset yang terkumpul terdiri dari variabel usia, jenis kelamin, tensi darah, indeks massa tubuh, kadar gula darah acak, riwayat hipertensi, riwayat kardiovaskular, kurangnya aktivitas.

Tabel 3. 3 Probabilitas prior

P(DM/Non DM)	54%	46%	100%
--------------	-----	-----	------

Langkah kedua adalah menghitung probabilitas prior. Tabel 3.3 menunjukkan persentase probabilitas pasien yang berisiko DM sebanyak 54%, sedangkan persentase non DM sebanyak 46%.

Tabel 3. 4 Probabilitas Kondisional

	DM	Non DM
P(U ...)		
Dewasa Awal	0%	8%
Dewasa Akhir	0%	17%
Lansia Awal	29%	17%
Lansia Akhir	50%	33%
Manula	21%	25%
	100%	100%
P(JK ...)	DM	Non DM
Laki-laki	43%	17%
Perempuan	57%	83%
	100%	100%
P(TENSI ...)	DM	Non DM
Normal	43%	0%
Pra-hipertensi	14%	25%
HT tingkat I	21%	42%
HT tingkat II	21%	33%
	100%	100%
P(IMT ...)	DM	Non DM
Kurus	0%	0%
Normal	71%	50%
Gemuk	29%	50%
	100%	100%
P(GDA ...)	DM	Non DM
Normal	50%	83%
Berisiko	50%	17%
	100%	100%
P(HT ...)	DM	Non DM
Ya	43%	50%
Tidak	57%	50%
	100%	100%
P(K ...)	DM	Non DM
Ya	21%	0%
Tidak	79%	100%
	100%	100%
P(A ...)	DM	Non DM
Cukup	79%	75%
Kurang	21%	25%
	100%	100%

Langkah ketiga adalah menghitung probabilitas kondisional berdasarkan variabel. Tabel 3.4 menunjukkan persentase probabilitas pasien yang berisiko DM dan non DM dari variabel usia, jenis kelamin, tensi darah, indeks massa tubuh, kadar gula darah acak, riwayat hipertensi, riwayat kardiovaskular, kurangnya aktivitas.

Tabel 3. 5 Data *testing*

DATA TESTING	Umur	Jenis Kelam	TENSI	IMT	GDA	RIW HT	RIW KAF	AKTIFIT	DIAGNOSA
LANSIA AWAL	Perempuan	HT TINGKAT I	GEMUK	BERISIKO	YA	TIDAK	CUKUP	DM	
MANULA	Perempuan	PRA-HIPERTENSI	KURUS	BERISIKO	YA	TIDAK	CUKUP	DM	
LANSIA AWAL	Laki-laki	HT TINGKAT I	GEMUK	BERISIKO	TIDAK	TIDAK	CUKUP	DM	
LANSIA AWAL	Laki-laki	PRA-HIPERTENSI	NORMAL	BERISIKO	TIDAK	TIDAK	CUKUP	DM	
MANULA	Perempuan	PRA-HIPERTENSI	NORMAL	BERISIKO	TIDAK	TIDAK	CUKUP	DM	
LANSIA AWAL	Laki-laki	NORMAL	GEMUK	BERISIKO	TIDAK	TIDAK	CUKUP	DM	
LANSIA AWAL	Perempuan	HT TINGKAT I	NORMAL	NORMAL	TIDAK	TIDAK	CUKUP	NON DM	
LANSIA AKHIR	Perempuan	HT TINGKAT II	NORMAL	BERISIKO	YA	TIDAK	CUKUP	NON DM	
MANULA	Laki-laki	PRA-HIPERTENSI	NORMAL	BERISIKO	TIDAK	TIDAK	CUKUP	NON DM	
LANSIA AKHIR	Perempuan	NORMAL	GEMUK	NORMAL	YA	TIDAK	KURANG	NON DM	

Langkah keempat adalah menginputkan dataset baru untuk *ditesting*.

Tabel 3.5 merupakan dataset baru yang akan diprediksi.

Tabel 3. 6 Probabilitas posterior

PREDICTION	Class Prediction	DM	Non DM
DM	0.00071203	0.00061031	
Non DM	0.00000000	0.00000000	
DM	0.00071203	0.00061031	
DM	0.00118671	0.00101718	
DM	0.00118671	0.00101718	
DM	0.00142405	0.00122062	
DM	0.00237342	0.00203436	
DM	0.00311512	0.00267010	
DM	0.00089003	0.00076289	
DM	0.00067966	0.00058257	

Langkah kelima adalah menghitung probabilitas posterior. Tabel 3.6 merupakan hasil perhitungan menggunakan rumus *Naïve Bayes*. Hasil perhitungan akan dibandingkan jika DM lebih besar daripada Non DM maka hasilnya positif DM.

Tabel 3. 7 Hasil prediksi

Predicted	Class	
	DM	Non DM
DM	5	4
Non DM	1	0

Langkah keenam adalah menambahkan atau membuat kolom untuk output prediksi. Tabel 3.7 merupakan hasil prediksi dari perhitungan *Naïve Bayes*.

BAB 4. HASIL DAN PEMBAHASAN

4.1 Mengumpulkan data dan *pre-processing* data untuk kebutuhan sistem deteksi dini penyakit *Diabetes Mellitus (Communication)*.

Tahap ini menjelaskan proses pengumpulan dan persiapan data yang menjadi fondasi utama dalam pengembangan sistem deteksi dini *diabetes mellitus*. Data yang digunakan merupakan data historis pasien yang diambil dari Rumah Sakit Citra Husada. Proses ini sangat krusial untuk memastikan kualitas dan validitas data sebelum digunakan dalam model *Naïve Bayes*.

4.1.1 Pengumpulan Data

Proses pengumpulan data merupakan langkah fundamental yang menentukan kualitas dan validitas hasil penelitian. Data yang digunakan adalah data rekam medis historis yang diperoleh dari Rumah Sakit Citra Husada. Pengambilan data dilakukan secara cermat melalui observasi dan dokumentasi rekam medis yang ada, dengan menyesuaikan standar (Depkes RI, 2008). Pendekatan ini memungkinkan peneliti untuk bekerja dengan data nyata yang merepresentasikan kondisi pasien sesungguhnya.

Tabel 4. 1 Data Mentah Pasien Berisiko DM dan Tidak Berisiko DM

No	Umur	Jenis Kelamin	TENSI	IMT	GDA	RIWAYAT HIPER-TENSI	RIWAYAT KARDIO-VASKULAR	AKTIVITAS	KLS
1	65.62	Laki-laki	150/90	29.3	185	YA	YA	CUKUP	BERISIKO
2	64.75	Perempuan	131/81	31.11	90	TIDAK	TIDAK	CUKUP	BERISIKO
3	65.75	Perempuan	131/82	32.11	91	TIDAK	TIDAK	CUKUP	BERISIKO
4	59.12	Laki-laki	130/80	23.44	98	TIDAK	TIDAK	CUKUP	BERISIKO
5	55.32	Perempuan	170/90	25	107	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
6	51.24	Laki-laki	143/92	34.23	416	TIDAK	YA	CUKUP	BERISIKO
7	35.56	Laki-laki	90/70	18.52	75	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
8	80.99	Perempuan	130/72	26.67	176	YA	TIDAK	CUKUP	BERISIKO
9	64.06	Laki-laki	139/76	19.53	227	TIDAK	TIDAK	-	BERISIKO
10	67.20	Perempuan	113/60	29.14	290	TIDAK	TIDAK	CUKUP	BERISIKO
...	
596	68.30	Perempuan	186/100	27	164	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
597	65.02	Perempuan	132/87	27	106	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
598	63.59	Perempuan	132/78	22	138	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO

No	Umur	Jenis Kelamin	TENSI	IMT	GDA	RIWAYAT HIPER-TENSI	RIWAYAT KARDIO-VASKULAR	AKTIVITAS	KLS
599	64.96	Laki-laki	169/77	22	308	YA	TIDAK	CUKUP	TIDAK BERISIKO
600	50.48	Perempuan	140/90	27	134	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO

Tautan data lengkap : <https://s.id/riIMP>

Tabel 4.1 menunjukkan total 600 data rekam medis pasien berhasil dikumpulkan, yang terdiri dari 300 data pasien dengan diagnosis Diabetes Mellitus (DM) dan 300 data pasien dengan diagnosis Hipertensi (HT). Data pasien Berisiko DM berfungsi sebagai kelas positif yang menjadi target utama sistem untuk diidentifikasi. Sementara itu, data pasien HT digunakan sebagai diagnosa pembanding, mewakili kelas Tidak Berisiko. Pemilihan data pasien hipertensi sebagai pembanding sangat relevan karena penyakit ini memiliki beberapa gejala dan faktor risiko yang tumpang tindih dengan diabetes, seperti usia, berat badan, dan gaya hidup (Ihsan & Adisasmita, 2023).

Setiap data pasien, baik dari penderita *diabetes mellitus* maupun hipertensi, memuat delapan variabel kunci yang ditetapkan sebagai fitur untuk klasifikasi. Variabel-variabel ini dipilih berdasarkan relevansinya dengan diagnosis penyakit dan ketersediaannya pada rekam medis di Rumah Sakit Citra Husada. Variabel-variabel tersebut meliputi Usia, Jenis Kelamin, Indeks Massa Tubuh (IMT), Tensi Darah, Kadar Gula Darah Acak (GDA), Riwayat Hipertensi, Riwayat Kardiovaskuler, dan Kurangnya Aktivitas. Pentingnya pemeriksaan variabel tersebut dalam mendeteksi risiko Diabetes Melitus juga terbukti dalam penelitian Delfina et al. (2021) yang menunjukkan bahwa usia, IMT, dan aktivitas fisik merupakan faktor risiko signifikan terhadap kejadian DM Tipe 2 pada populasi Indonesia.

4.1.2 *Pre-processing* Data

Pre-processing data merupakan tahap penting yang dilakukan untuk memastikan data yang digunakan dalam sistem deteksi dini *diabetes mellitus* memiliki kualitas tinggi dan siap diolah oleh model *Naïve Bayes*. Proses ini melibatkan serangkaian langkah sistematis yang bertujuan untuk memperbaiki data dari segala bentuk ketidak sempurnaan, seperti nilai yang hilang, format yang tidak konsisten, atau duplikasi. Tahap *pre-processing* ini terdiri dari empat

langkah utama, yaitu *data cleaning*, *data integration*, *data selection*, dan *data transformation*.

Pada tahap *pre-processing* data, data mentah yang telah dikumpulkan dikonversi menjadi format yang dapat diproses oleh algoritma *Naïve Bayes*. Seperti yang terlihat pada tabel, data mentah awal (Tabel 4.2) memiliki nilai numerik seperti usia (contohnya, 65.62), tensi (150/90), dan IMT (29.3). Untuk mempermudah proses klasifikasi dan analisis, data-data numerik ini diubah menjadi kategori atau label.

Tabel 4. 2 Data Kategorikal

No	Umur	Jenis Kelamin	TENSI	IMT	GDA	RIWAYAT HIPERTENSI	RIWAYAT KARDIO-VASKULAR	AKTIVITAS	KLS
1	LANSIA AKHIR	LAKI-LAKI	HIPERTENSI I	GEMUK	NORMAL	YA	YA	CUKUP	BERISIKO
2	LANSIA AKHIR	PEREMPUAN	PRA-HIPERTENSI	GEMUK	NORMAL	TIDAK	TIDAK	CUKUP	BERISIKO
3	LANSIA AKHIR	PEREMPUAN	PRA-HIPERTENSI	GEMUK	NORMAL	TIDAK	TIDAK	CUKUP	BERISIKO
4	LANSIA AKHIR	LAKI-LAKI	PRA-HIPERTENSI	NORMAL	NORMAL	TIDAK	TIDAK	CUKUP	BERISIKO
5	LANSIA AWAL	PEREMPUAN	HIPERTENSI II	NORMAL	NORMAL	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
6	LANSIA AWAL	LAKI-LAKI	HIPERTENSI I	GEMUK	BERISIKO	TIDAK	YA	CUKUP	BERISIKO
7	DEWASA AWAL	LAKI-LAKI	NORMAL	NORMAL	NORMAL	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
8	MANULA	PEREMPUAN	PRA-HIPERTENSI	GEMUK	NORMAL	YA	TIDAK	CUKUP	BERISIKO
9	LANSIA AKHIR	LAKI-LAKI	PRA-HIPERTENSI	NORMAL	BERISIKO	TIDAK	TIDAK	CUKUP	BERISIKO
10	MANULA	PEREMPUAN	NORMAL	GEMUK	BERISIKO	TIDAK	TIDAK	CUKUP	BERISIKO
...
596	MANULA	PEREMPUAN	HIPERTENSI II	GEMUK	NORMAL	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
597	LANSIA AKHIR	PEREMPUAN	PRA-HIPERTENSI	GEMUK	NORMAL	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
598	LANSIA AKHIR	PEREMPUAN	PRA-HIPERTENSI	NORMAL	NORMAL	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
599	LANSIA AKHIR	LAKI-LAKI	HIPERTENSI II	NORMAL	BERISIKO	YA	TIDAK	CUKUP	TIDAK BERISIKO
600	LANSIA AWAL	PEREMPUAN	HIPERTENSI I	GEMUK	NORMAL	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO

Tautan data lengkap : <https://s.id/riIMP>

Proses kategorisasi ini menghasilkan data yang lebih terstruktur (Tabel 4.2). Contohnya, nilai usia numerik dikategorikan menjadi rentang usia seperti LANSIA AKHIR atau MANULA. Tensi darah yang semula berupa angka (150/90) diubah menjadi kategori HIPERTENSI. Hal yang sama juga dilakukan pada IMT (29.3), yang dikategorikan menjadi GEMUK. Sementara itu, GDA

(Glukosa Darah Acak) dikategorikan menjadi NORMAL atau BERISIKO berdasarkan rentang nilai yang telah ditentukan.

Tujuan dari proses kategorisasi ini adalah untuk menyederhanakan data, mengurangi kompleksitas, dan memastikan bahwa setiap variabel memiliki nilai diskrit yang sesuai dengan kebutuhan algoritma *Naïve Bayes*. Dengan demikian, data menjadi lebih bersih, konsisten, dan siap untuk digunakan dalam tahap pelatihan dan pengujian model.

a. *Data Cleaning* (Pembersihan Data)

Langkah pertama dalam *pre-processing* adalah *Data Cleaning*. Pada tahap ini, data rekam medis pasien dari Rumah Sakit Citra Husada diperiksa untuk mengidentifikasi dan menangani berbagai anomali. Fokus utama adalah pada penanganan nilai yang hilang (*missing values*). Contohnya, pada variabel Kurangnya Aktivitas, sering ditemukan nilai kosong yang secara medis tidak valid. Nilai-nilai ini tidak dapat dihapus karena akan mengurangi jumlah data yang signifikan, sehingga ditangani dengan metode imputasi menggunakan nilai rata-rata (*mean*) dari masing-masing variabel. Dengan cara ini, distribusi data dapat dipertahankan, dan model dapat dilatih dengan data yang lebih lengkap.

Tabel 4. 3 Tabel kategorikal

No	Umur	Jenis Kelamin	TENSI	IMT	GDA	RIWAYAT HIPERTENSI	RIWAYAT KARDIO-VASKULAR	AKTIVITAS	KLS
1	LANSIA AKHIR	LAKI-LAKI	HIPERTENSI I	GEMUK	NORMAL	YA	YA	CUKUP	BERISIKO
2	LANSIA AKHIR	PEREMPUAN	PRA-HIPERTENSI	GEMUK	NORMAL	TIDAK	TIDAK	CUKUP	BERISIKO
3	LANSIA AKHIR	PEREMPUAN	PRA-HIPERTENSI	GEMUK	NORMAL	TIDAK	TIDAK	CUKUP	BERISIKO
...
11	MANULA	PEREMPUAN	NORMAL	GEMUK	BERISIKO	TIDAK	TIDAK	-	BERISIKO
...
255	LANSIA AKHIR	LAKI-LAKI	PRA-HIPERTENSI	NORMAL	NORMAL	TIDAK	TIDAK	-	BERISIKO
...
320	DEWASA AWAL	PEREMPUAN	HIPERTENSI II	GEMUK	-	YA	TIDAK	CUKUP	TIDAK BERISIKO
...
296	MANULA	PEREMPUAN	HIPERTENSI II	GEMUK	NORMAL	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
297	LANSIA AKHIR	PEREMPUAN	PRA-HIPERTENSI	GEMUK	NORMAL	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO

No	Umur	Jenis Kelamin	TENSI	IMT	GDA	RIWAYAT HIPER-TENSI	RIWAYAT KARDIO-VASKULAR	AKTIVITAS	KLS
298	LANSIA AKHIR	PEREMPUAN	PRA-HIPERTENSI	NORMAL	NORMAL	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO
299	LANSIA AKHIR	LAKI-LAKI	HIPERTENSI II	NORMAL	BERISIKO	YA	TIDAK	CUKUP	TIDAK BERISIKO
300	LANSIA AWAL	PEREMPUAN	HIPERTENSI I	GEMUK	NORMAL	TIDAK	TIDAK	CUKUP	TIDAK BERISIKO

Tautan data lengkap : <https://s.id/riIMP>

Pada tabel data, terlihat adanya simbol tanda hubung (-) pada kolom AKTIVITAS (baris 11 dan 255) dan GDA (baris 320). Simbol ini mengindikasikan bahwa data pada entri tersebut hilang atau tidak terisi. Dilakukan teknik imputasi dengan mengisi nilai yang hilang tersebut menggunakan nilai rata-rata (*mean*). Oleh karena itu, semua nilai "-" pada kolom tersebut akan diganti menjadi "CUKUP". Proses ini memastikan bahwa semua entri memiliki data yang lengkap dan konsisten, sehingga model dapat memprosesnya tanpa kendala. Pentingnya data lengkap (penanganan *missing value*) sebagai bagian dari *pre-processing* juga sejalan dengan penelitian Widianti & Pratama (2024) yang menemukan bahwa banyaknya data kosong ('*missing values*') secara signifikan mengurangi akurasi model prediksi, dan mengisi data hilang sebelum pelatihan model sangat meningkatkan performa.

Tahap *pre-processing* data juga mencakup penanganan data duplikat. Meskipun tidak terlihat jelas dalam cuplikan tabel, jika ditemukan adanya data pasien yang terdaftar lebih dari satu kali, salah satu data duplikat tersebut akan dihapus. Langkah ini dilakukan untuk memastikan setiap data yang digunakan dalam model bersifat unik dan valid, sehingga tidak menyebabkan bias dalam proses pelatihan model. Dengan demikian, data yang dihasilkan dari tahap pembersihan ini menjadi lebih bersih, lengkap, dan siap untuk tahap *pre-processing* selanjutnya.

b. *Data Selection* (Pemilihan Data)

Tahap *Data Selection* berfokus pada pemilihan fitur-fitur yang paling relevan untuk proses klasifikasi. Berdasarkan studi literatur terkait, dipilih delapan variabel kunci dari rekam medis yang dianggap paling berpengaruh terhadap diagnosis diabetes. Variabel-variabel tersebut

meliputi Usia, Jenis Kelamin, Indeks Massa Tubuh (IMT), Tensi Darah, Kadar Gula Darah Acak (GDA), Riwayat Hipertensi, Riwayat Kardiovaskuler, dan Kurangnya Aktivitas. Pemilihan fitur yang tepat sangat penting untuk meningkatkan akurasi dan efisiensi model *Naïve Bayes*.

Pada tahap seleksi data (*Data Selection*), prosesnya tidak lagi berfokus pada penentuan variabel atau kelas target karena hal tersebut telah dilakukan pada fase awal penelitian. Sebaliknya, tahap ini memastikan bahwa data yang telah dikumpulkan, yaitu 300 data pasien Diabetes Mellitus (DM) dan 300 data pasien Hipertensi (HT), sudah benar-benar sesuai dan siap untuk digunakan. Seleksi ini memastikan tidak ada data yang anomali atau tidak relevan yang ikut terproses. Data pasien DM dikelompokkan secara spesifik sebagai kelas positif, sementara data pasien HT digunakan sebagai kelompok pembanding yang merepresentasikan kelas non-DM.

c. *Data Transformation* (Transformasi Data)

Tahap *Data Transformation* (Transformasi Data) merupakan proses penting dalam *pre-processing* yang bertujuan untuk mengubah data ke dalam format yang sesuai dengan kebutuhan algoritma *Naïve Bayes*. Proses ini mengubah semua data kategorikal menjadi numerik.

Tabel 4. 4 Tabel numerical

No	Umur	Jenis Kelamin	Tensi	IMT	GDA	Riwayat Hipertensi	Riwayat Kardiovaskular	Aktivitas	Kls
1	3	1	2	2	0	0	0	1	Berisiko
2	3	0	1	2	0	1	1	1	Berisiko
3	3	0	1	2	0	1	1	1	Berisiko
4	3	1	1	1	0	1	1	1	Berisiko
5	2	0	3	1	0	1	1	1	Tidak Berisiko
6	2	1	2	2	1	1	0	1	Berisiko
7	0	1	0	1	0	1	1	1	Tidak Berisiko
8	4	0	1	2	0	0	1	1	Berisiko
9	3	1	1	1	1	1	1	1	Berisiko
10	4	0	0	2	1	1	1	1	Berisiko
...	
596	4	0	3	2	0	1	1	1	Tidak Berisiko
597	3	0	1	2	0	1	1	1	Tidak Berisiko

No	Umur	Jenis Kelamin	Tensi	IMT	GDA	Riwayat Hipertensi	Riwayat Kardiovaskular	Aktivitas	Kls
598	3	0	1	1	0	1	1	1	Tidak Berisiko
599	3	1	3	1	1	0	1	1	Tidak Berisiko
600	2	0	2	2	0	1	1	1	Tidak Berisiko

Tautan data lengkap : <https://s.id/riIMP>

Langkah selanjutnya adalah mengubahnya menjadi bentuk numerik setelah seluruh data berada dalam bentuk kategorikal (baik dari hasil diskretisasi maupun data asli seperti "Jenis Kelamin" dan "Riwayat Hipertensi"). Proses ini dikenal sebagai *encoding*, di mana setiap kategori diwakili oleh sebuah angka unik (contohnya, "Laki-laki" diubah menjadi 0 dan "Perempuan" menjadi 1). Penelitian (Id, 2021) mengatakan bahwa untuk memudahkan algoritma *machine learning* dalam melakukan prediksi salah satunya pendekatan *one-hot encoding* dengan membuat suatu kolom baru dari variabel kategorikal, di mana setiap kategori menjadi kolom baru dengan nilai 0 atau 1 yang di mana nilai 0 menandakan tidak mewakili ada dan 1 mewakili ada. Seluruh dataset berada dalam format numerik yang terstruktur, bersih, dan siap untuk diproses oleh algoritma *Naïve Bayes* untuk membangun model klasifikasi.

4.1.3 Analisis Deskriptif Data Pasien

Bagian ini menyajikan analisis deskriptif terhadap data pasien yang dikumpulkan dalam penelitian ini. Data disajikan dalam bentuk tabel yang memisahkan masing-masing faktor risiko, seperti usia, jenis kelamin, Indeks Massa Tubuh (IMT), tekanan darah, kadar gula darah, riwayat hipertensi, riwayat kardiovaskular, dan kurangnya aktivitas. Analisis ini bertujuan untuk memberikan gambaran umum mengenai karakteristik demografi dan klinis dari populasi sampel, yang menjadi dasar untuk pembahasan lebih mendalam pada bagian selanjutnya. Data yang disajikan dalam tabel ini akan dianalisis lebih lanjut untuk mengidentifikasi pola dan hubungan yang relevan, serta untuk membandingkan temuan penelitian ini dengan studi-studi terdahulu.

Tabel 4. 5 Tabel rekapitulasi data pasien

Variabel	Kategori	Jumlah data
Umur	Dewasa awal	17
	Dewasa akhir	60

Variabel	Kategori	Jumlah data
Jenis kelamin	Lansia awal	174
	Lansia akhir	215
	Manula	134
IMT	Laki-laki	206
	Perempuan	394
Tekanan darah	Kurus	16
	Normal	360
	Gemuk	224
GDA	Normal	106
	Pra-hipertensi	135
	Hipertensi I	179
	Hipertensi II	180
Riwayat HT	Normal	203
	Berisiko	397
Riwayat Kardiovaskular	Ya	220
	Tidak	380
Kurangnya Aktivitas	Ya	42
	Tidak	558
Kurangnya Aktivitas	Cukup	544
	Kurang	56

Analisis profil pasien menunjukkan konsistensi yang kuat dengan temuan epidemiologi sebelumnya mengenai faktor risiko diabetes melitus. Dari total 600 pasien, sebanyak 523 di antaranya berada dalam kelompok usia lansia awal, lansia akhir, dan manula yang mana masuk dalam kelompok usia dalam berbagai studi telah diidentifikasi sebagai populasi dengan prevalensi tertinggi untuk diabetes. Hal ini sejalan dengan hasil Riskesdas 2018 yang menunjukkan bahwa prevalensi diabetes meningkat secara signifikan pada kelompok usia ≥ 55 tahun, menjadikan usia sebagai faktor risiko non-modifikasi yang paling dominan dalam perkembangan penyakit ini (Kemenkes RI, 2018). Selain itu, dominasi pasien perempuan (394 dari 600) juga relevan dengan penelitian yang menunjukkan bahwa wanita *pasca-menopause* memiliki peningkatan risiko diabetes akibat perubahan hormonal yang memengaruhi metabolisme glukosa dan distribusi lemak tubuh. Studi oleh (Freeman, Acevedo, & Pennings., 2023) menjelaskan bahwa resistensi insulin merupakan mekanisme kunci yang berperan dalam perkembangan diabetes tipe 2, terutama pada individu dengan obesitas, penuaan, dan perubahan hormonal seperti yang terjadi pada wanita

pasca-menopause. Dari segi klinis, sebanyak 224 pasien memiliki Indeks Massa Tubuh (IMT) dalam kategori gemuk, sementara mayoritas lainnya berada pada kategori *overweight*. Obesitas telah dikenal sebagai salah satu faktor risiko utama diabetes tipe 2, karena berkontribusi pada resistensi insulin dan disfungsi sel beta pankreas (Paleva, 2019). Kombinasi dengan temuan tekanan darah tidak normal (pra-hipertensi hingga hipertensi II) pada sebagian besar pasien memperkuat keterkaitan antara sindrom metabolik dan diabetes.

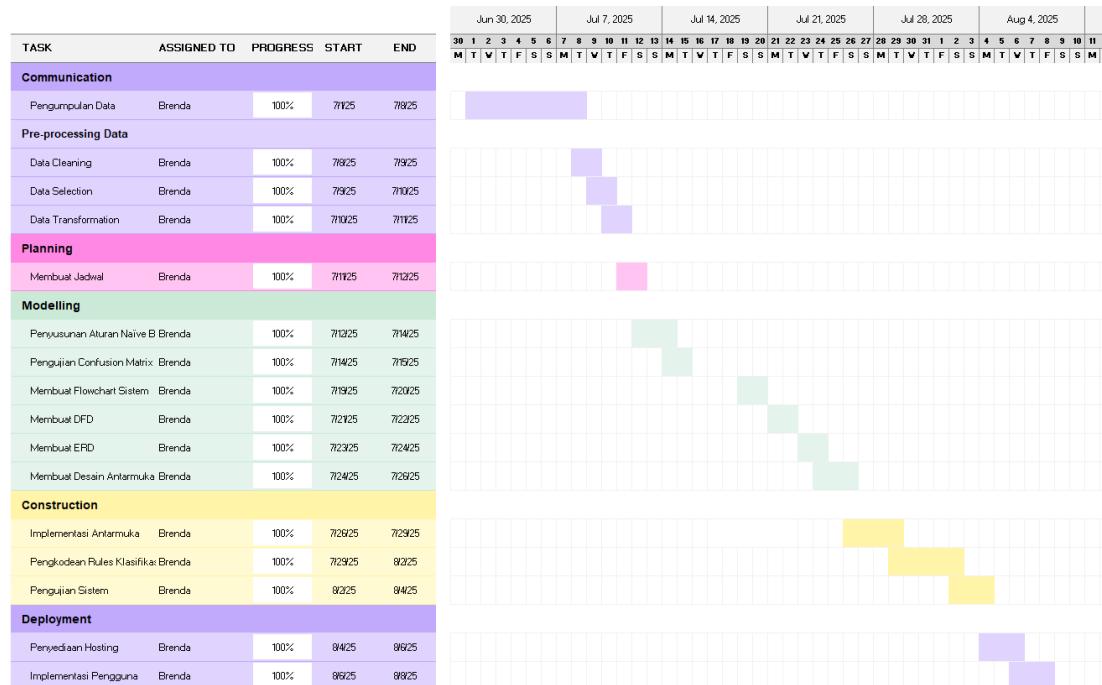
Fakta bahwa 397 pasien memiliki kadar gula darah berisiko secara langsung menunjukkan bahwa populasi ini sangat dekat dengan atau telah mengalami onset diabetes, dan memperkuat urgensi intervensi dini. Riwayat hipertensi (220 pasien) dan penyakit kardiovaskular (42 pasien) juga mendukung bukti komorbiditas umum pada penderita diabetes, sebagaimana dijelaskan oleh International Diabetes Federation (IDF, 2021), yang mencatat bahwa penderita diabetes memiliki risiko dua kali lipat lebih tinggi untuk mengalami komplikasi kardiovaskular. Meskipun 544 pasien dilaporkan memiliki aktivitas fisik yang cukup, keberadaan 56 pasien yang kurang aktif secara fisik menunjukkan bahwa gaya hidup tetap menjadi faktor kontribusi yang penting, terutama karena inaktivitas fisik adalah faktor risiko modifikasi yang sangat berpengaruh terhadap kontrol glukosa darah.

4.2 Membuat perencanaan proses pengembangan sistem deteksi dini penyakit *Diabetes Mellitus* (*Planning*).

Tahap perencanaan adalah langkah fundamental dalam model pengembangan *Waterfall*. Pada tahap ini, disusun sebuah rencana kerja yang komprehensif untuk seluruh proses pengembangan sistem. Rencana ini menjadi panduan utama untuk memastikan setiap aktivitas, mulai dari perancangan hingga pengujian, dapat berjalan secara terstruktur dan sesuai jadwal. Perencanaan ini mencakup penetapan aktivitas-aktivitas utama, estimasi durasi, serta alokasi sumber daya. Tujuannya adalah untuk mengelola proyek secara efektif, meminimalisir risiko, dan memastikan sistem deteksi dini diabetes dapat diselesaikan dengan tepat waktu. Manfaat penggunaan metode penjadwalan dalam proyek nyata telah diteliti oleh Lengkong (2023) yang menjelaskan

bahwa penggunaan *Microsoft Project* untuk membuat *Gantt Chart* dan lintasan kritis untuk penjadwalan memberikan gambaran jelas mengenai urutan aktivitas dan durasi, sehingga proyek kesehatan tersebut bisa selesai secara optimal dan sesuai waktu rencana.

Gantt Chart adalah sebuah alat visualisasi yang digunakan untuk merencanakan, mengelola, dan memantau kemajuan proyek. Berdasarkan gambar diatas, *Gantt Chart* ini memetakan seluruh tahapan penelitian dan pengembangan sistem deteksi dini *diabetes mellitus*. Setiap baris (*task*) merepresentasikan sebuah aktivitas, seperti "Pengumpulan Data," "Pre-processing Data," "Modelling," hingga "Deployment." Rentang waktu dari setiap aktivitas digambarkan dengan sebuah bar, yang menunjukkan durasi dan jadwal pengerjaan.



Gambar 4. 3 Gantt chart waterfall

Gantt Chart di atas menjadi panduan utama untuk memastikan semua tahapan berjalan sesuai rencana. Sebagai contoh, dapat dilihat bahwa setelah "Pre-processing Data" selesai, dilanjutkan dengan tahap "Modelling," di mana aktivitas seperti "Penyusunan Aturan Naïve Bayes" dan "Pengujian Confusion Matrix" dilakukan secara berurutan. Kemudian, setelah model selesai,

dilanjutkan dengan tahap "*Construction*" yang mencakup "Implementasi Antarmuka" dan "Pengujian Sistem," hingga akhirnya tahap "Deployment" yang merupakan penutup dari seluruh proyek. Dengan adanya *Gantt Chart* ini, memudahkan untuk melacak jadwal, memonitor kemajuan, dan memastikan proyek selesai tepat waktu.

4.3 Menyusun dan menguji algoritma serta menganalisis dan mendesain sistem deteksi dini penyakit *Diabetes Mellitus (Modelling)*.

4.3.1 Pengujian Kinerja Algoritma *Naïve Bayes*

a. Pengujian Kinerja menggunakan *tools RapidMiner*

Proses *mining* dilakukan dengan menggunakan *tools RapidMiner* untuk mengevaluasi kinerja model *Naïve Bayes*. Pengujian ini dilakukan dengan membandingkan beberapa komposisi data *training* dan data *testing* menggunakan teknik pemilihan sampel *automatic sampling* pada total dataset sebanyak 600 data. Komposisi yang diuji meliputi 95:5, 90:10, 85:15, 80:20, 75:25, 70:30, 65:35, 60:40, 55:45, dan 50:50. Teknik *automatic sampling* ini memastikan bahwa model dievaluasi secara acak dengan berbagai skenario pembagian data. Penelitian terdahulu mendukung penggunaan skenario proporsi pembagian data berganda sebagai cara untuk memperoleh estimasi performa model yang lebih robust. Misalnya, sebuah studi tahun 2024 menemukan bahwa rasio *training/test* yang bervariasi (termasuk 90:10, 80:20, 70:30, 60:40, 50:50) secara signifikan memengaruhi metrik seperti akurasi dan generalisasi model, menunjukkan bahwa proporsi data pelatihan yang lebih besar belum tentu selalu memberikan performa yang jauh lebih baik jika tidak diuji dengan beberapa rasio berbeda (Bichri et al., 2024).

Sebagai contoh, jika dataset penelitian ini terdiri dari 600 data yang terbagi menjadi data kelas "berisiko" dan "tidak berisiko", dengan pembagian 70:30, maka model akan dilatih dengan 70% dari total data (420 data) dan diuji dengan 30% sisanya (180 data). Proses ini diulang untuk setiap komposisi yang diuji, memastikan bahwa evaluasi model komprehensif dan tidak bias. Accuracy tertinggi dari pengujian ini (73,81%)

ditemukan pada komposisi 65:35, yang menunjukkan konfigurasi optimal untuk model ini dalam mengklasifikasikan risiko.

Tabel 4. 6 Hasil akurasi model

Perbandingan Komposisi Dataset	Akurasi
95:5	73,43%
90:10	70%
85:15	68,89%
80:20	69,17%
75:25	71,33%
70:30	71,67%
65:35	73,81%
60:40	73,75%
55:45	73,7%
50:50	69,67%

Hasil ini menunjukkan bahwa alokasi data latih sebanyak 65% sudah memadai bagi model untuk mempelajari pola dari data, sementara 35% data uji cukup representatif untuk mengevaluasi kinerja model. Walaupun komposisi 55:45 juga menunjukkan akurasi yang tinggi (73,7%), komposisi 65:35 tetap menjadi yang paling optimal. Akurasi ini menjadi dasar penting untuk menunjukkan bahwa model Naïve Bayes memiliki kemampuan yang baik dalam mengklasifikasikan risiko. Hal ini sejalan dengan penelitian Mandalika et al. (2024) yang menunjukkan bahwa model *Naïve Bayes* dengan *k-fold cross-validation* dapat mencapai akurasi tertinggi sebesar 73% pada data uji. Penelitian ini menekankan pentingnya pemilihan fitur yang tepat untuk meningkatkan performa model dalam prediksi diabetes.

b. Pengujian Kinerja Menggunakan *Confusion Matrix*

Berdasarkan penghitungan kinerja pada proses *mining*, diperoleh akurasi tertinggi dalam pembuatan aturan klasifikasi dengan perbandingan komposisi data *training* dan data *testing* sebesar 65:35. Tabel 4.2 berikut ini merupakan perbandingan hasil klasifikasi dengan diagnosis riil.

Tabel 4. 7 Perbandingan hasil klasifikasi dengan data aktual

No	No RM	Diagnosis	Hasil Prediksi
1	*****9	Berisiko	Berisiko
2	*****9	Berisiko	Berisiko
3	*****4	Tidak Berisiko	Tidak Berisiko
4	*****3	Berisiko	Berisiko
5	*****4	Tidak Berisiko	Berisiko

No	No RM	Diagnosis	Hasil Prediksi
6	*****1	Berisiko	Berisiko
7	*****2	Berisiko	Berisiko
8	*****1	Berisiko	Berisiko
9	*****2	Berisiko	Berisiko
10	*****6	Tidak Berisiko	Tidak Berisiko
...
201	*****2	Berisiko	Berisiko
202	*****1	Tidak Berisiko	Berisiko
203	*****1	Berisiko	Berisiko
204	*****1	Berisiko	Berisiko
205	*****3	Berisiko	Berisiko
206	*****9	Tidak Berisiko	Tidak Berisiko
207	*****2	Berisiko	Berisiko
208	*****2	Tidak Berisiko	Berisiko
209	*****2	Berisiko	Berisiko
300	*****2	Berisiko	Berisiko

Tautan data lengkap : <https://s.id/riIMP>

Tabel perbandingan data riil dan hasil prediksi ini menyajikan sampel data yang digunakan untuk pengujian model. Setiap baris menunjukkan perbandingan antara diagnosis aktual (Diagnosis) dengan hasil prediksi dari model (Hasil Prediksi). Hasil ini memberikan gambaran langsung mengenai akurasi model pada setiap data uji. Contohnya, pada baris ke-202, diagnosis riil adalah "Tidak Berisiko" tetapi model salah memprediksi sebagai "Berisiko". Sebaliknya, pada baris ke-203, model berhasil memprediksi "Berisiko" dengan benar. Analisis per baris seperti ini sangat penting untuk mengidentifikasi pola kesalahan yang mungkin terjadi, seperti ketika model cenderung salah mengklasifikasikan data "Tidak Berisiko" menjadi "Berisiko". Selain menunjukkan nilai akurasi, tabel Confusion Matrix juga dapat menunjukkan nilai precision, dan recall. Hasil Confusion Matrix disajikan dalam Tabel 4.3 berikut.

Tabel 4. 8 Confusion matrix

		Prediksi Berisiko DM	Prediksi Tidak Berisiko DM
Kelas Aktual	Berisiko DM	TP=90	FN=26
	Tidak Berisiko DM	FP=36	TN=58

Confusion matrix ini adalah ringkasan dari hasil prediksi keseluruhan, yang memberikan gambaran yang lebih terstruktur tentang kinerja model. Matriks ini memisahkan hasil prediksi menjadi empat

kuadran. *True Positive* (TP) menunjukkan 90 data "Berisiko" yang berhasil diprediksi dengan benar. *False Negative* (FN) menunjukkan 26 data "Berisiko" yang salah diprediksi sebagai "Tidak Berisiko". *False Positive* (FP) menunjukkan 36 data "Tidak Berisiko" yang salah diprediksi sebagai "Berisiko". Terakhir, *True Negative* (TN) menunjukkan 58 data "Tidak Berisiko" yang berhasil diprediksi dengan benar. Matriks ini sangat berguna untuk menghitung *accuracy*, *precision*, dan *recall*. Hasil hitungan sebagai berikut.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / \text{Total Data} = (90 + 7) / 210 = 70,5\%$$

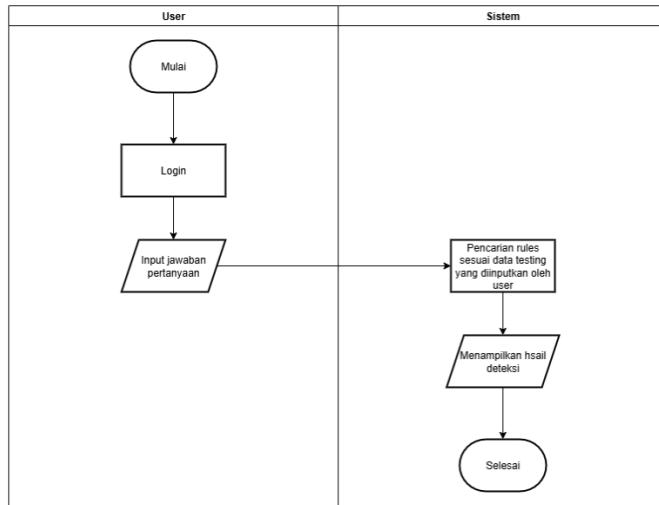
$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) = 90 / (90 + 36) = 71,4\%$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) = 90 / (90 + 26) = 77,6\%$$

Berdasarkan nilai-nilai dari confusion matrix, dihitung tiga metrik evaluasi utama untuk mengukur kinerja model secara kuantitatif. *Accuracy* dihitung sebesar 70,5%, yang berarti 70,5% dari total data uji berhasil diklasifikasikan dengan benar. *Precision* sebesar 71,4% menunjukkan bahwa dari semua prediksi "Berisiko", 71,4% di antaranya adalah benar. Nilai *precision* ini sangat penting untuk kasus di mana kesalahan prediksi positif (pasien sehat diprediksi sakit) harus dihindari. Terakhir, *Recall* sebesar 77,6% menunjukkan bahwa model mampu mengidentifikasi 77,6% dari seluruh kasus "Berisiko" yang ada. Hal ini sejalan dengan penelitian Salsabila et al. (2024) yang menunjukkan bahwa model Naïve Bayes dapat mencapai nilai recall sebesar 60,8%. Meskipun nilai recall ini menunjukkan kemampuan model dalam mengidentifikasi kasus positif, nilai tersebut masih dapat ditingkatkan untuk meningkatkan sensitivitas model dalam mendeteksi risiko diabetes. Kombinasi ketiga *matrix* ini memberikan gambaran yang lengkap dan seimbang mengenai seberapa andal model dalam memprediksi risiko diabetes.

4.3.2 Analisis sistem

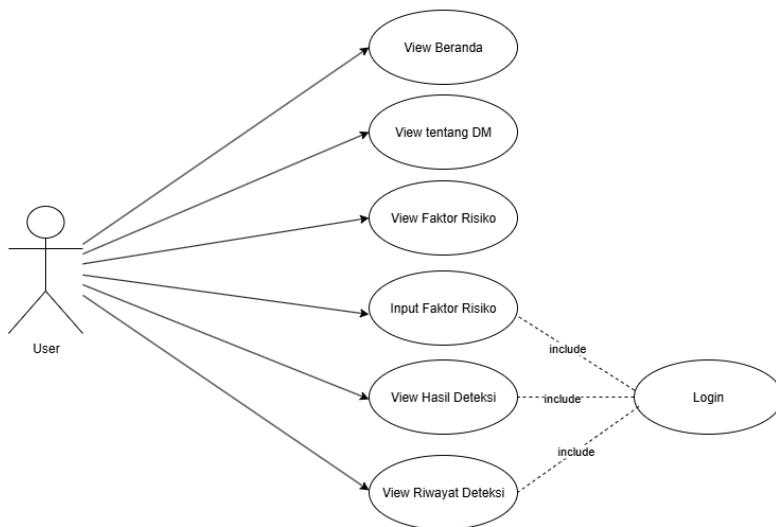
a. *Flowchart* Sistem



Gambar 4. 4 Flowchart Sistem

Secara keseluruhan, alur kerja sistem deteksi dini diabetes mellitus ini dapat dijelaskan melalui flowchart pada Gambar 4.2. Alur dimulai ketika pengguna mengakses sistem dan melakukan proses login. Setelah berhasil masuk, pengguna dapat memulai sesi deteksi dengan memasukkan jawaban dari serangkaian pertanyaan yang diajukan oleh sistem. Pertanyaan-pertanyaan ini berfungsi sebagai data masukan (testing data) yang berisi variabel-variabel kunci seperti usia, IMT, dan kondisi kesehatan lainnya. Setelah pengguna menyelesaikan input data, sistem akan memproses data tersebut. Proses inti di sini adalah pencarian aturan (*rules*) yang sesuai dengan data masukan pengguna. Aturan-aturan ini sebelumnya telah dibangun menggunakan algoritma *Naïve Bayes* pada tahap pelatihan model. Sistem kemudian mencocokkan data pengguna dengan aturan-aturan ini untuk menghitung probabilitas kemungkinan pengguna mengidap atau tidak mengidap *diabetes milletus*. Hasil dari perhitungan ini kemudian ditampilkan sebagai hasil deteksi. Proses ini memungkinkan pengguna mendapatkan deteksi yang cepat dan berbasis data. Alur kerja ini diakhiri dengan tampilan hasil dan pengguna dapat keluar dari sistem.

b. *Use Case* dan *Activity Diagram*



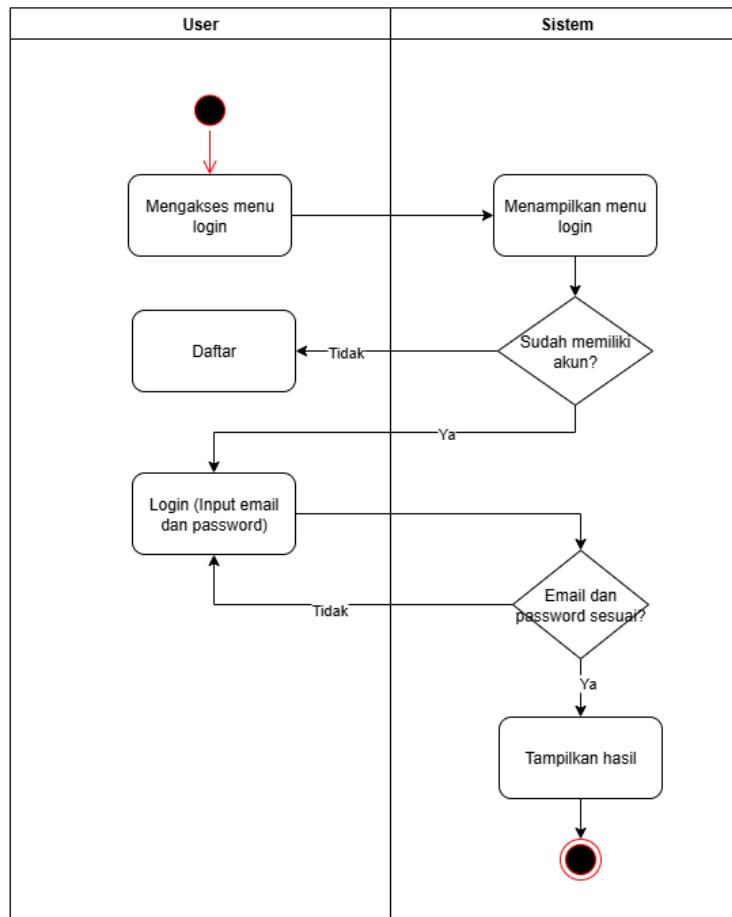
Gambar 4. 5 *Use case diagram*

Gambar 4.3 menunjukkan *use case diagram* memvisualisasikan fungsionalitas utama yang dapat diakses oleh pengguna. Sebagai sebuah sistem deteksi dini diabetes, alur utama dimulai dengan pengguna yang dapat melihat *View Beranda*, *View tentang DM*, dan *View Faktor Risiko*. Ketiga use case ini bersifat informatif, memberikan pengetahuan dasar mengenai diabetes dan faktor-faktor risikonya kepada pengguna.

Proses deteksi dimulai dari pengguna harus melakukan *Login*. Proses ini sangat penting karena terkait dengan dua fungsionalitas utama lainnya, yaitu *View Hasil Deteksi* dan *View Riwayat Deteksi*. Keduanya terhubung dengan *use case Login* melalui hubungan *include*. Ini menandakan bahwa pengguna harus sudah dalam kondisi *login* untuk dapat melihat riwayat deteksi mereka.

Pengguna dapat melakukan *Input Faktor Risiko* setelah *Login*, di mana mereka memasukkan data-data yang dibutuhkan oleh sistem untuk dianalisis. Data ini kemudian diproses untuk menghasilkan *View Hasil Deteksi*. Proses ini mencerminkan alur utama sistem, di mana masukan dari pengguna diolah untuk memberikan hasil prediksi. Semua *use case*

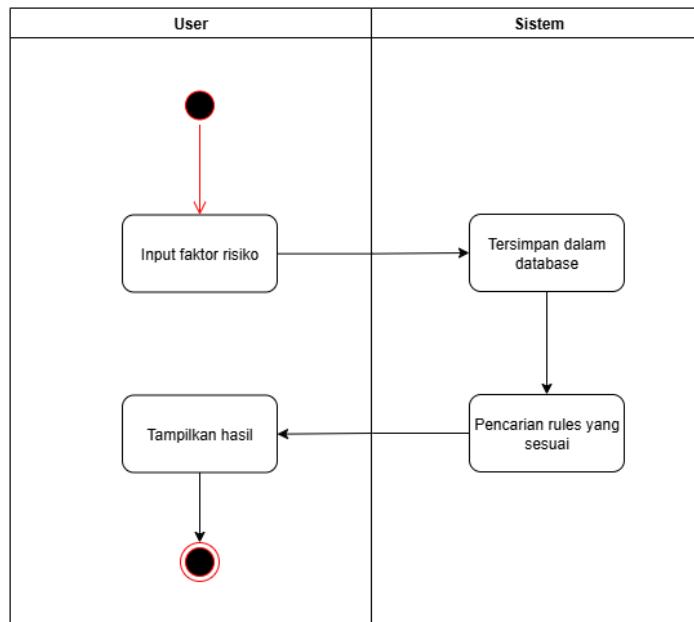
ini saling berhubungan dan dirancang untuk memberikan pengalaman yang lancar dan terstruktur bagi pengguna, dari sekedar mendapatkan informasi hingga melakukan deteksi dan melihat riwayat hasil.



Gambar 4. 6 Use case diagram user

Gambar 4.4 merupakan *Activity Diagram* untuk proses *login* yang menggambarkan interaksi antara pengguna dan sistem. Alur dimulai dari pengguna yang mengakses halaman *login*. Sistem akan menampilkan formulir *login* dan memvalidasi keberadaan akun pengguna. Jika pengguna belum memiliki akun, alur akan mengarah ke proses pendaftaran. Sebaliknya jika pengguna sudah terdaftar, mereka akan diminta untuk memasukkan *email* dan *password*. Sistem kemudian memverifikasi kredensial tersebut. Apabila *email* dan *password* sesuai, pengguna akan berhasil masuk ke sistem. Namun, jika tidak sesuai,

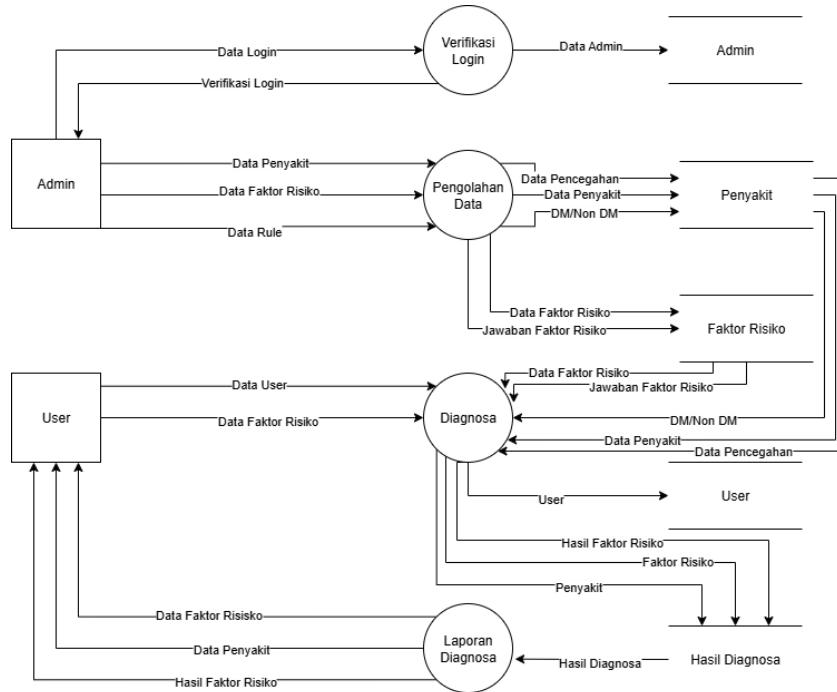
pengguna akan tetap berada di halaman *login* untuk mencoba kembali. Proses ini memastikan bahwa hanya pengguna yang terverifikasi yang dapat mengakses fitur-fitur utama sistem.



Gambar 4. 7 Activity diagram sistem deteksi dini

Gambar 4.5 merupakan *Activity Diagram* untuk proses deteksi dini diabetes yang menggambarkan serangkaian aktivitas yang dimulai setelah pengguna berhasil *login*. Pengguna akan melakukan input data faktor risiko yang relevan, seperti usia, kadar glukosa, dan IMT. Data yang dimasukkan oleh pengguna ini kemudian akan diproses oleh sistem. Langkah pertama adalah penyimpanan data tersebut ke dalam basis data, berfungsi sebagai riwayat deteksi bagi pengguna. Selanjutnya, sistem akan melakukan proses inti, yaitu pencarian aturan (*rules*) yang sesuai dengan data yang diinputkan pengguna. Aturan ini adalah hasil dari pelatihan model *Naïve Bayes*. Setelah sistem menemukan dan menerapkan aturan yang relevan, hasil prediksi akan ditampilkan kepada pengguna. Proses ini dirancang untuk memberikan informasi yang cepat dan akurat mengenai risiko diabetes berdasarkan data yang diberikan oleh pengguna.

c. *Data Flow Diagram (DFD)*



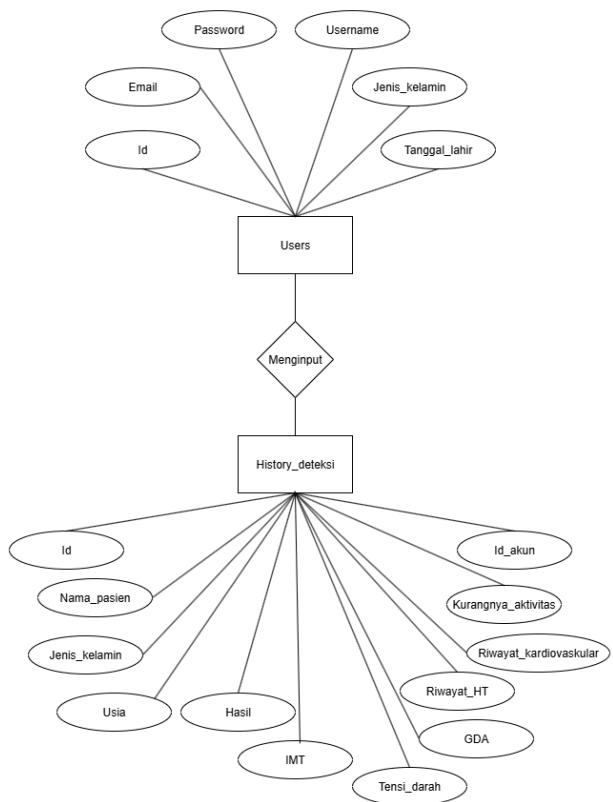
Gambar 4.5 merupakan Data Flow Diagram (DFD) Level 0 memvisualisasikan sistem deteksi dini diabetes secara umum. Diagram ini menunjukkan tiga entitas utama yang berinteraksi dengan sistem, yaitu *Admin*, *User*, dan Laporan Diagnosa. Sistem itu sendiri diwakili oleh tiga proses utama: Verifikasi *Login*, Pengolahan Data, dan Diagnosa.

Alur data dimulai dari entitas *Admin* yang memberikan data *login*, data penyakit, data faktor risiko, dan data aturan (*rule*) ke dalam sistem. Data ini digunakan untuk memverifikasi *login admin* dan memelihara basis data yang digunakan oleh sistem. Proses Pengolahan Data menerima masukan dari admin dan menghasilkan data pencegahan, data penyakit, serta data Berisiko/Tidak Berisiko yang akan digunakan dalam proses diagnosa.

Sementara itu, entitas *User* memasukkan data pengguna dan data faktor risiko ke dalam proses Diagnosa. Proses ini akan mengolah data dari pengguna dengan basis data yang ada untuk menghasilkan jawaban

faktor risiko, serta data Berisiko/Tidak Berisiko, data penyakit, dan data pencegahan yang kemudian dikirimkan kembali ke pengguna sebagai hasil deteksi. Data dari proses diagnosa juga mengalir ke entitas Laporan Diagnosa untuk menghasilkan laporan berupa data faktor risiko, data penyakit, dan hasil faktor risiko. Diagram ini secara keseluruhan menggambarkan bagaimana data mengalir dari entitas eksternal, diproses di dalam sistem, dan dikembalikan sebagai informasi yang

d. *Entity Relationship Diagram (ERD)*



Gambar 4. 8 *Entity-Relationship Diagram*

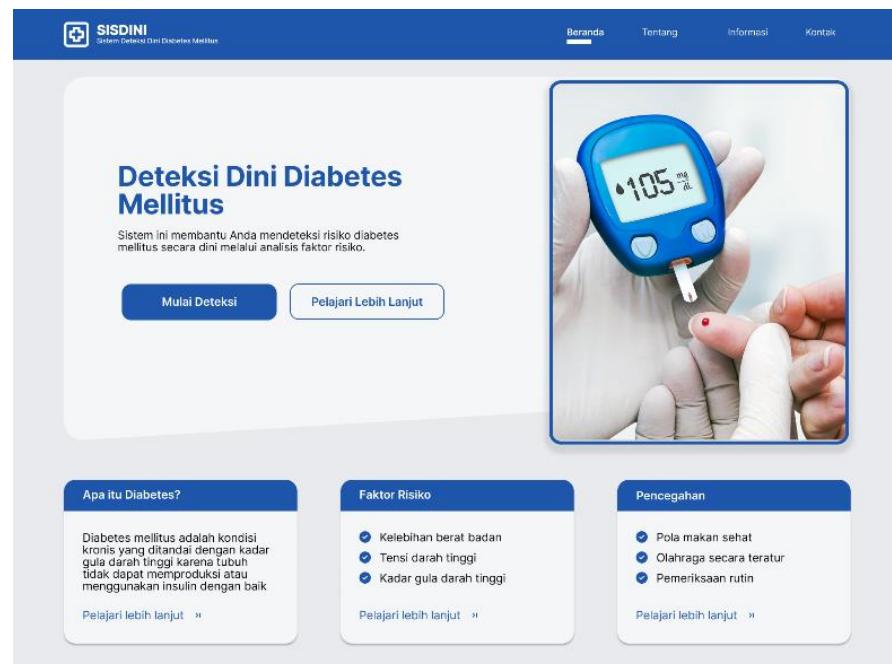
Gambar 4.6 merupakan *Entity-Relationship Diagram* (ERD) ini memvisualisasikan struktur basis data sistem deteksi dini diabetes. Diagram ini menunjukkan dua entitas utama *Users* dan *History_deteksi*. Entitas *Users* berfungsi untuk menyimpan data identitas pengguna. Atribut-atribut yang terkait dengan entitas ini meliputi *Id*, *Email*,

Password, Username, Jenis_kelamin, dan Tanggal_lahir. Atribut Id berperan sebagai kunci utama (*primary key*) untuk mengidentifikasi setiap pengguna secara unik.

Hubungan antara kedua entitas ini digambarkan oleh entitas relasional Menginput. Relasi ini menunjukkan bahwa satu pengguna (*Users*) dapat menginputkan (menginput) data deteksi beberapa kali, yang akan disimpan dalam entitas History_deteksi. Dengan demikian, ini membentuk hubungan *one-to-many* antara *Users* dan History_deteksi.

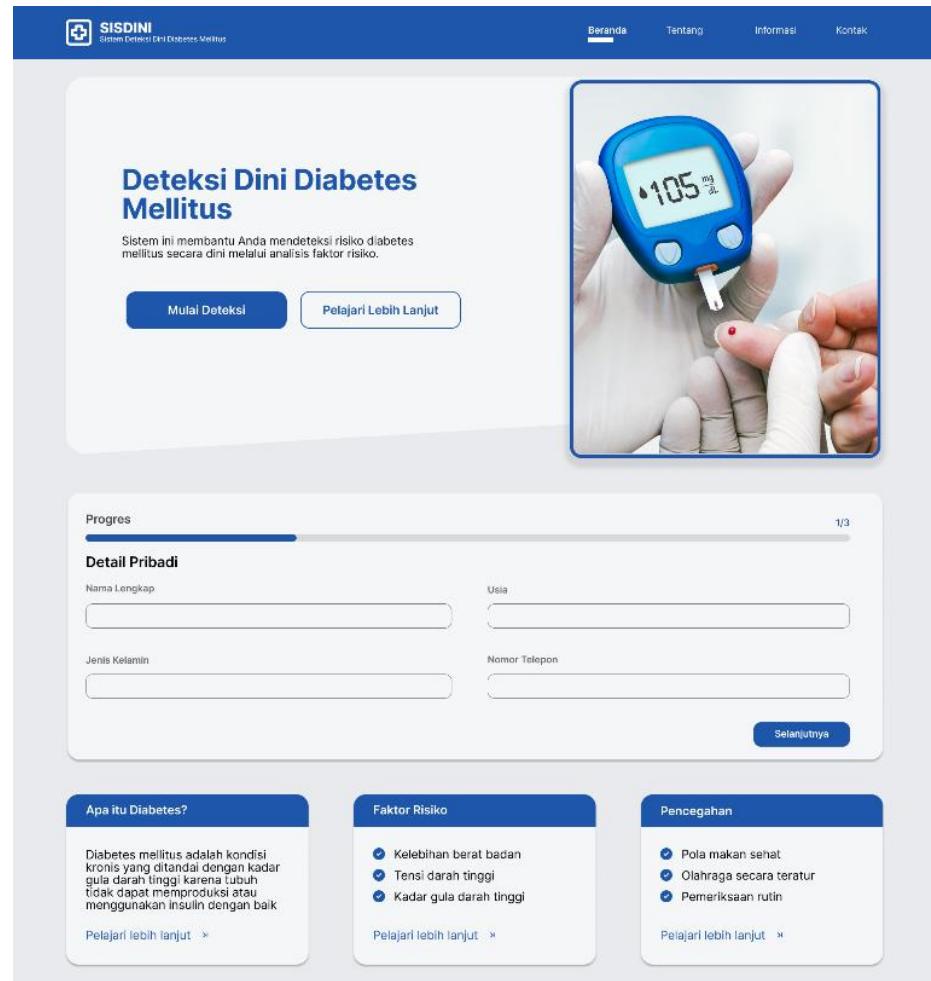
Entitas History_deteksi berfungsi untuk menyimpan riwayat hasil deteksi dari setiap pengguna. Atribut-atribut yang tercakup dalam entitas ini sangat beragam, mulai dari data pribadi pasien seperti Id, Nama_pasien, Jenis_kelamin, Usia, IMT, dan Tensi_darah, hingga data hasil diagnosis seperti Hasil dan faktor-faktor risiko lainnya (Id_akun, Kurangnya_aktivitas, Riwayat_kardiovaskular, dan GDA). Atribut Id_akun pada entitas History_deteksi berfungsi sebagai kunci asing (*foreign key*) yang merujuk pada Id di entitas *Users*, sehingga setiap riwayat deteksi dapat dihubungkan kembali ke pengguna yang bersangkutan. Struktur basis data ini memastikan data tersimpan secara terstruktur dan terintegrasi dengan baik.bermanfaat.

4.3.3 Desain Antarmuka Sistem



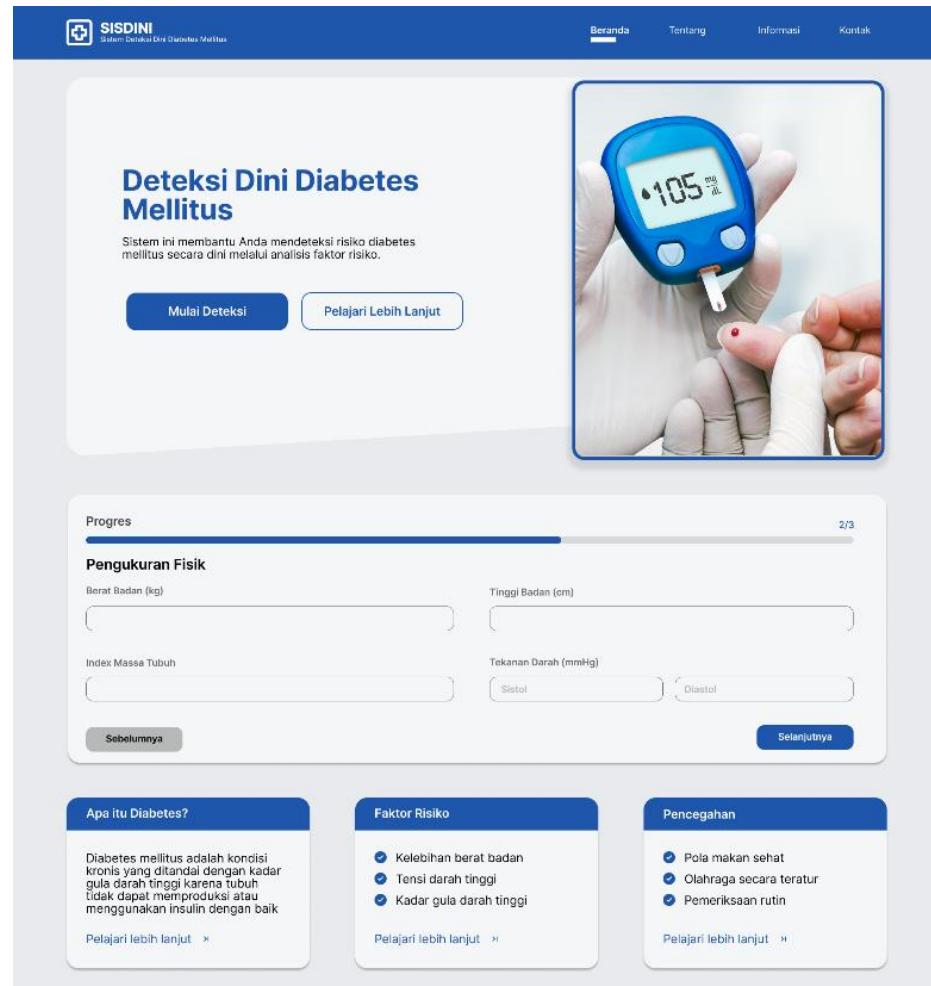
Gambar 4. 9 Tampilan Home

Gambar di atas menunjukkan tampilan home Sistem Deteksi *Dini Diabetes Mellitus*, yang dirancang untuk memungkinkan pengguna memantau status kesehatan mereka dengan mudah dan menerima informasi pencegahan diabetes secara lebih efektif. Antarmuka yang mudah digunakan membuat sistem ini memungkinkan pengguna untuk dengan mudah mengakses informasi kesehatan dan memahami tindakan yang dapat diambil untuk mencegah risiko diabetes.



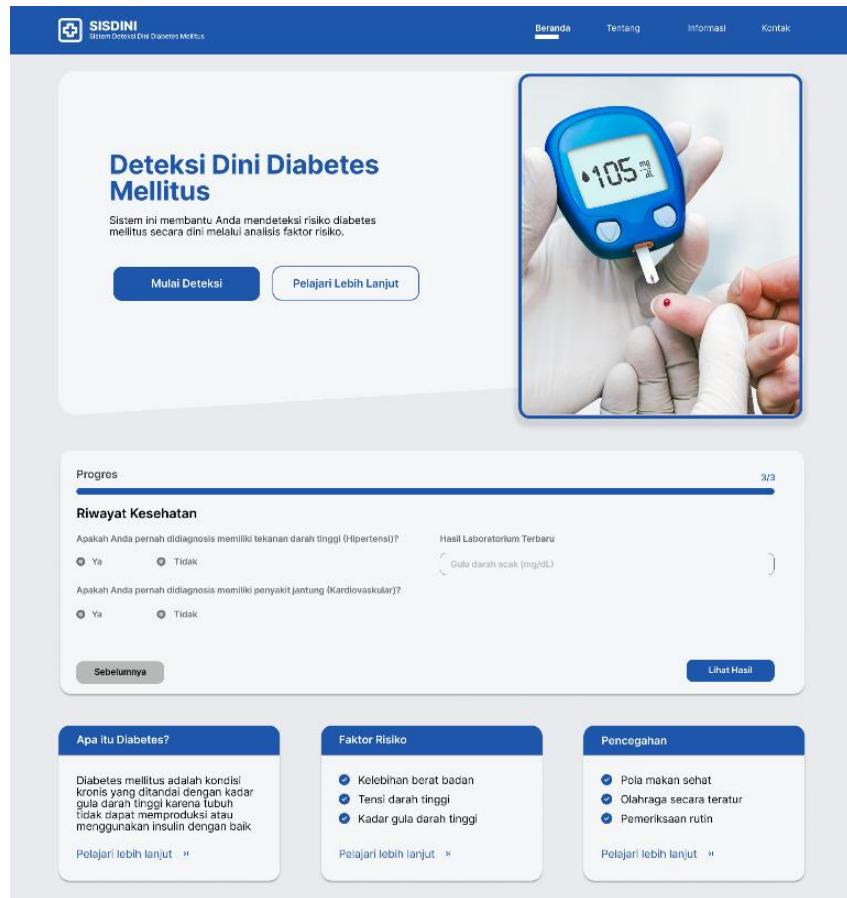
Gambar 4. 10 Tampilan Pengisian Data Pasien

Gambar di atas menunjukkan tampilan pengisian data pasien pada sistem deteksi dini diabetes mellitus yang memungkinkan tenaga kesehatan dan pengguna untuk memasukkan identitas pasien dan data-data penting untuk memastikan analisis yang lebih akurat dan tepat sasaran. Fitur ini dilengkapi dengan nama, usia, dan jenis kelamin pasien.



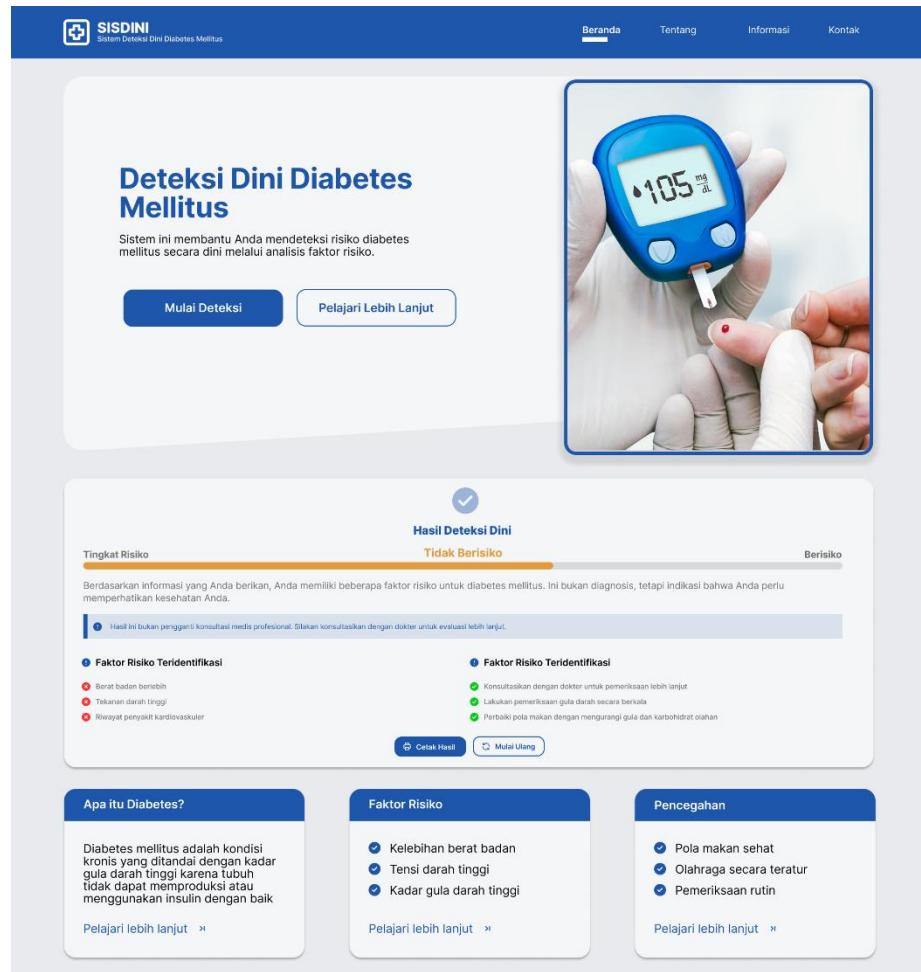
Gambar 4. 11 Tampilan Pemeriksaan Fisik

Gambar di atas menunjukkan tampilan hasil tes pasien dalam pemeriksaan fisik, yang berfungsi sebagai tempat untuk merekam hasil tes, indeks massa tubuh dan tensi darah, sehingga lebih mudah untuk menilai kondisi pasien. Hasil tes yang direkam dapat digunakan untuk memantau kesehatan pasien secara teratur dan membantu staf medis menentukan pengobatan yang tepat.



Gambar 4. 12 Tampilan Pengisian Riwayat Penyakit

Gambar di atas menunjukkan tampilan pengisian riwayat pasien pada sistem deteksi dini diabetes mellitus yang secara sistematis mencatat riwayat medis pasien, termasuk gaya hidup dan riwayat penyakit untuk meningkatkan keakuratan skrining. Informasi yang terekam dapat digunakan untuk memberikan saran gaya hidup sehat kepada pasien yang disesuaikan dengan kondisi mereka, sehingga mendukung upaya pencegahan.



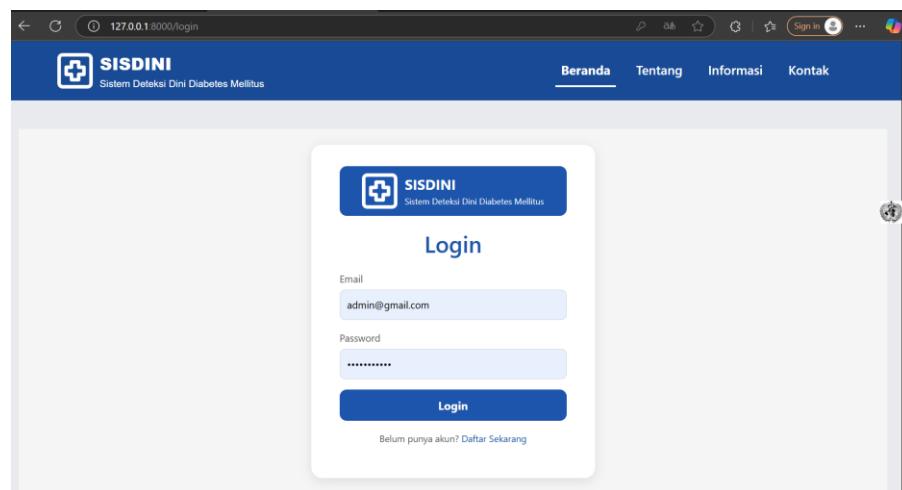
Gambar 4. 13 Tampilan Hasil Deteksi Dini

Gambar di atas menunjukkan tampilan hasil deteksi dini pasien pada sistem deteksi dini diabetes mellitus yang menampilkan hasil analisis berdasarkan data yang dimasukkan, memberikan saran mengenai langkah selanjutnya yang harus dilakukan untuk menjaga kesehatan pasien dengan lebih baik. Sistem ini tidak hanya memberikan hasil deteksi, tetapi juga rekomendasi berdasarkan data yang dianalisa sehingga pasien mengetahui tindakan apa yang harus dilakukan selanjutnya.

4.4 Melakukan pembangunan sistem deteksi dini penyakit *Diabetes Mellitus* (*Construction*).

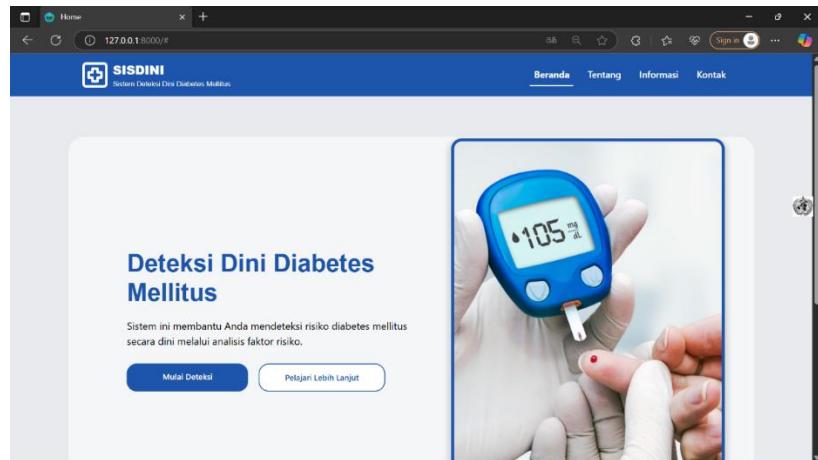
4.4.1 Implementasi Antarmuka

Secara umum, implementasi antarmuka ini mencakup beberapa komponen kunci. Pertama, tampilan halaman utama yang menjadi gerbang awal bagi pengguna. Kedua, formulir atau form yang dirancang untuk mengumpulkan data dari pengguna, seperti data faktor risiko. Ketiga, tampilan hasil deteksi yang menyajikan informasi prediksi secara jelas dan mudah dipahami. Terakhir, tampilan riwayat deteksi yang memungkinkan pengguna melihat kembali hasil-hasil sebelumnya. Setiap antarmuka dirancang dengan memperhatikan prinsip-prinsip *user experience* (UX) dan *user interface* (UI) yang baik, memastikan alur kerja yang logis dan pengalaman pengguna yang lancar.



Gambar 4. 14 Halaman login dokter dan admin

Halaman *Login* merupakan pintu utama bagi pengguna (dokter dan admin) untuk dapat mengakses sistem SISDINI (Sistem Deteksi Dini Diabetes Mellitus). Pada halaman ini disediakan *form* autentikasi yang terdiri dari dua field input, yaitu *email* dan *password*, yang digunakan untuk memverifikasi identitas pengguna. Pada area utama halaman, terdapat tombol *Login* sebagai aksi utama untuk masuk ke dalam sistem setelah pengguna mengisi data yang diperlukan.



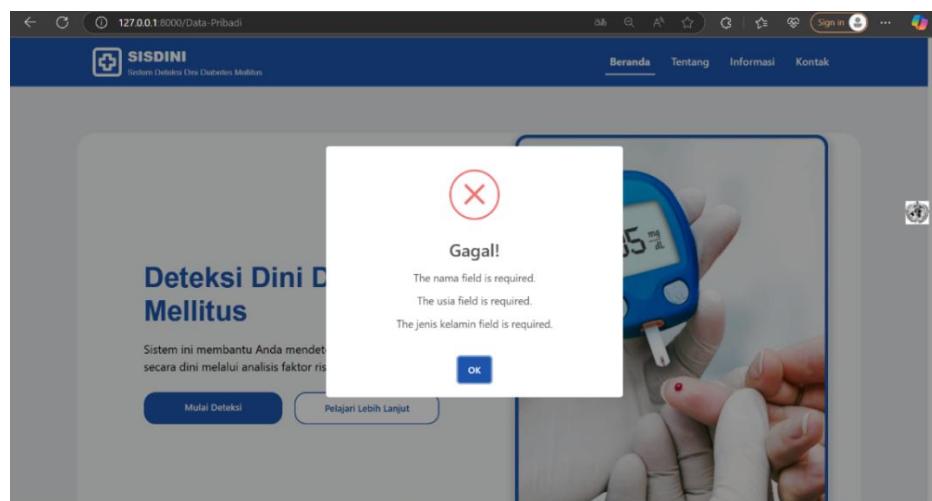
Gambar 4. 15 Halaman utama pasien

Halaman utama berfungsi sebagai gerbang utama sistem yang menyambut pengguna (pasien) dan memberikan gambaran singkat tentang fungsi aplikasi. Seperti yang terlihat pada gambar, halaman ini dirancang dengan tata letak yang bersih dan modern. Tombol "Mulai Deteksi" adalah interaksi utama yang akan membawa pengguna ke halaman pengisian data. Tampilan ini juga dilengkapi dengan logo aplikasi, navigasi (*navbar*) yang jelas, dan ilustrasi yang relevan untuk memperkuat visualisasi dari tujuan sistem.

Gambar 4. 16 Halaman deteksi

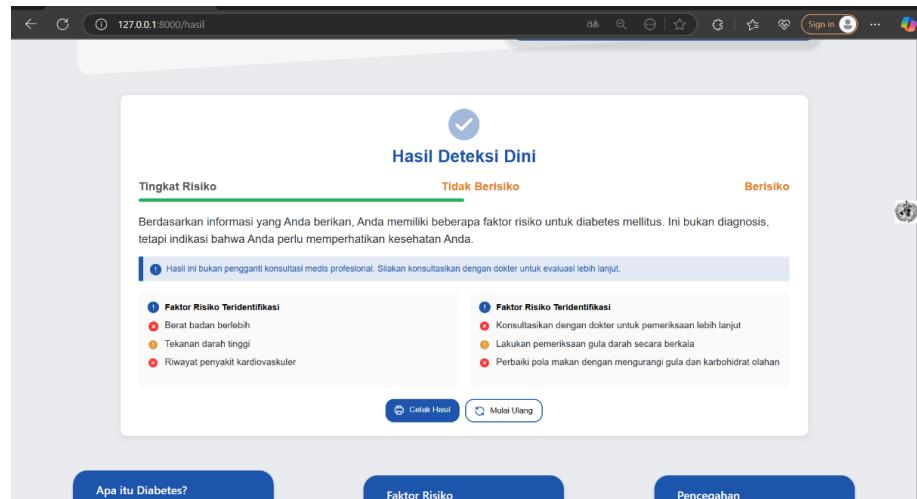
Halaman ini merupakan inti dari sistem deteksi, di mana pengguna diminta untuk memasukkan data pribadi dan faktor risiko. Tampilan ini dirancang dengan formulir (form) yang terstruktur dan dibagi menjadi beberapa

bagian, seperti "Detail Pribadi" dan "Faktor Risiko". Pada bagian "Detail Pribadi", pengguna mengisi informasi dasar seperti nama, usia, jenis kelamin, dan nomor telepon. Selanjutnya, pengguna akan diarahkan untuk memasukkan data-data medis yang menjadi faktor risiko utama. Formulir ini dirancang dengan navigasi yang intuitif, seperti tombol "Selanjutnya", dan dilengkapi dengan progress bar di bagian atas untuk menunjukkan tahapan pengisian data.



Gambar 4. 17 *Pop-up gagal*

Fitur ini berfungsi untuk memeriksa setiap *field* pada formulir, dan jika ada *field* yang kosong atau tidak diisi sesuai format yang benar, sistem akan menampilkan pesan kesalahan. Seperti yang terlihat pada Gambar 4. ketika pengguna menekan tombol "Selanjutnya" tanpa mengisi *field* yang wajib diisi, sebuah jendela (*pop-up*) akan muncul. Jendela ini menampilkan pesan Gagal! dan detail kesalahan, seperti The name field is required, The usia field is required, dan The jenis kelamin field is required. Ini adalah langkah penting dalam menjamin kualitas data yang masuk ke dalam sistem, yang pada akhirnya akan mempengaruhi keakuratan hasil prediksi.



Gambar 4. 18 Halaman hasil deteksi

Halaman ini adalah output utama dari sistem, di mana hasil prediksi ditampilkan setelah data berhasil diproses. Seperti yang terlihat pada gambar, halaman ini dirancang untuk memberikan informasi yang jelas dan terstruktur. Tampilan utama menunjukkan hasil deteksi dalam bentuk kategori "Tingkat Risiko", yang diklasifikasikan menjadi "Tidak Berisiko" atau "Berisiko". Selain hasil prediksi, halaman ini juga menyajikan informasi edukatif dan rekomendasi. Terdapat bagian "Faktor Risiko Teridentifikasi" yang merinci faktor-faktor risiko pasien yang terdeteksi, serta "Faktor Risiko Teridentifikasi" yang berisi rekomendasi praktis, seperti anjuran untuk konsultasi lebih lanjut dengan dokter atau perbaikan pola makan.

4.4.2 Pengkodean Sistem

Pengkodean sistem, atau *system encoding*, adalah proses di mana informasi atau data diubah menjadi format digital yang dapat diproses dan disimpan oleh komputer. Proses ini sangat penting karena komputer hanya dapat memahami dan berinteraksi dengan data dalam bentuk biner (0 dan 1). Pengkodean sistem memastikan bahwa data, seperti teks, gambar, suara, atau video, dapat diterjemahkan secara konsisten dan akurat dari satu sistem ke sistem lain.

Sistem dimulai dengan pengumpulan data melalui formulir multi-langkah yang intuitif. Formulir ini merupakan inti dari halaman *web*, dirancang untuk memandu pengguna secara bertahap dalam memasukkan informasi yang

relevan. Data yang diinputkan akan dikirimkan ke *server* menggunakan metode POST menuju rute deteksi.store, di mana proses validasi akan dimulai.

Alur pengisian data dalam sistem ini dirancang dalam bentuk formulir interaktif tiga langkah. Pada Langkah 1, pengguna diminta untuk mengisi informasi pribadi dasar seperti nama, usia, jenis kelamin, dan nomor telepon. Setelah itu, formulir beralih ke Langkah 2, yang berfokus pada data pengukuran fisik. Pengguna memasukkan data aktivitas, tinggi, berat, dan tekanan darah. Fitur menarik pada tahap ini adalah perhitungan Indeks Massa Tubuh (IMT) yang dilakukan secara otomatis menggunakan JavaScript begitu pengguna memasukkan tinggi dan berat badan, memberikan hasil instan. Terakhir, Langkah 3 meminta pengguna untuk mengisi riwayat kesehatan, termasuk informasi mengenai hipertensi, penyakit kardiovaskular, dan hasil pemeriksaan gula darah. Setelah semua data terkumpul, pengguna dapat menekan tombol "Lihat Hasil" untuk mengirimkan data ke sistem, yang kemudian akan memproses dan menampilkan hasil deteksi risiko.

Setelah semua data diterima oleh DeteksiController.php, proses validasi dimulai untuk memastikan setiap input telah diisi dengan benar dan memiliki format yang sesuai.

```
public function store(Request $request)
{
    $validated = $request->validate([
        'nama'          => 'required|string|max:255',
        'usia'          => 'required|numeric',
        'jenis_kelamin' => 'required',
        'telepon'        => 'nullable|string',
        'aktivitas'      => 'nullable',
        'tinggi_badan'  => 'nullable|numeric',
        'berat_badan'   => 'nullable|numeric',
        'tekanan_darah' => 'nullable|string',
        'hipertensi'     => 'nullable',
        'gula_darah'     => 'nullable',
        'kardiovaskular' => 'nullable',
    ]);
}
```

Gambar 4. 19 Proses validasi data

Kode pada gambar 4.19 menjelaskan proses validasi dan normalisasi data yang diinputkan oleh pengguna sebelum data tersebut diolah lebih lanjut

oleh algoritma. Sistem dimulai saat pengguna mengisi formulir di aplikasi web. Data mentah seperti nama, usia, dan parameter kesehatan lainnya diterima oleh fungsi store(Request \$request) di DeteksiController.php. Kode ini pertama-tama akan memvalidasi data untuk memastikan kelengkapannya. Selanjutnya, data diproses melalui serangkaian langkah normalisasi dan *encoding*.

```
// =====
// Hitung IMT
// =====
$imt = null;
if (!empty($validated['tinggi_badan']) && !empty($validated['berat_badan'])) {
    $tinggi_m = $validated['tinggi_badan'] / 100;
    if ($tinggi_m > 0) {
        $imt = round($validated['berat_badan'] / ($tinggi_m * $tinggi_m), 2);
    }
}

// =====
// Parsing Tekanan Darah
// =====
$sistol = $diastol = null;
if (!empty($validated['tekanan_darah']) && strpos($validated['tekanan_darah'], '/') !== false) {
    [$sistol, $diastol] = array_map('trim', explode('/', $validated['tekanan_darah']));
    $sistol = is_numeric($sistol) ? (int) $sistol : null;
    $diastol = is_numeric($diastol) ? (int) $diastol : null;
}

// =====
// Konversi ke nilai kategorikal
// =====
$usia_asli = (int) $validated['usia'];
$usia_val = ($usia_asli >= 26 && $usia_asli <= 35) ? 0 :
            (($usia_asli >= 36 && $usia_asli <= 45) ? 1 :
            (($usia_asli >= 46 && $usia_asli <= 55) ? 2 :
            (($usia_asli >= 56 && $usia_asli <= 65) ? 3 : 4)));
$jkl_val = is_numeric($validated['jenis_kelamin'])
           ? (int) $validated['jenis_kelamin']
           : (strtolower($validated['jenis_kelamin']) == 'laki-laki' ? 1 : 0);
```

Gambar 4. 20 Proses normalisasi data

Proses awal dalam sistem ini adalah pra-pemrosesan data yang bertujuan mengubah data mentah dari pengguna menjadi format yang dapat diolah oleh algoritma. Langkah pertama adalah perhitungan IMT (Indeks Massa Tubuh), di mana berat dan tinggi badan digunakan untuk menghitung nilai IMT, yang kemudian dikelompokkan ke dalam kategori numerik: 0 untuk kurus, 1 untuk normal, dan 2 untuk gemuk/obesitas. Selanjutnya, sistem melakukan parsing tekanan darah; input seperti "120/80" diurai menjadi nilai sistolik dan diastolik, lalu dikategorikan menjadi nilai 0 hingga 3 sesuai dengan standar kesehatan. Terakhir, semua data kategorikal, seperti jenis kelamin dan aktivitas fisik, diubah menjadi nilai numerik melalui proses encoding variabel. Contohnya, "Laki-laki" dikodekan menjadi 1 dan

"Perempuan" menjadi 0. Seluruh proses ini memastikan bahwa semua data input dikonversi menjadi format yang konsisten dan terstruktur, menjadikannya siap untuk dianalisis oleh model *Naïve Bayes*.

```
// =====
// Baca Dataset Excel
// =====
$filePath    = public_path('data/matang.xlsx');
$spreadsheet = IOFactory::load($filePath);
$sheet       = $spreadsheet->getActiveSheet();
$rows        = $sheet->toArray();
array_shift($rows); // hapus header

$data_diabetes     = [];
$data_non_diabetes = [];
$exactMatchLabel   = null;

foreach ($rows as $row) {
    $record = [
        'usia'          => (int) $row[0],
        'jk'            => (int) $row[1],
        'imt'           => (int) $row[2],
        'td'            => (int) $row[3],
        'hipertensi'    => (int) $row[4],
        'kardiovaskular' => (int) $row[5],
        'aktivitas'     => (int) $row[6],
        'gd'            => (int) $row[7],
    ];
    $label = strtolower(trim($row[8]));
}
```

Gambar 4. 21 Proses analisis sistem

Setelah data pengguna selesai diproses, sistem mulai memuat dataset pelatihan dari file Excel yang berisi data rekam medis pasien terdahulu. Data ini sudah dilabeli sebagai "Berisiko" atau "Tidak Berisiko". Sistem akan mencoba mencocokkan data pengguna dengan data di dataset. Jika ditemukan kecocokan persis, sistem akan langsung menggunakan label yang sesuai dari dataset sebagai hasil prediksi. Namun, jika tidak ada kecocokan, sistem akan beralih ke algoritma Naïve Bayes.

```

// =====
// Tentukan hasil
// =====
if ($exactMatchLabel !== null) {
    $hasil = $exactMatchLabel;
} else {
    // cek faktor risiko manual
    $faktorRisiko = false;
    if ($gd_val == 1 || $td_val >= 2 || $imt_val == 2 || $hipertensi_val == 1 || $kardiovaskular_val == 1) {
        $faktorRisiko = true;
    }

    if ($faktorRisiko) {
        $hasil = "Berisiko";
    } else {
        // jika ingin tetap gunakan Naive Bayes
        $total = count($rows);
        if ($total > 0) {
            $p_diabetes = count($data_diabetes) / $total;
            $p_non_diabetes = count($data_non_diabetes) / $total;

            $likelihood = function ($data, $field, $value) {
                $total_data = count($data);
                if ($total_data == 0) return 1;

                $count = array_reduce($data, fn($carry, $item) =>
                    $carry + ($item[$field] == $value ? 1 : 0), 0);

                $categories = array_unique(array_column($data, $field));
                $num_categories = count($categories) ?: 1;

                return ($count + 1) / ($total_data + $num_categories);
            };
    }
}

```

Gambar 4. 22 Proses algoritma

Algoritma ini diawali dengan memisahkan dataset ke dalam dua kelompok: pasien yang dilabeli berisiko dan pasien yang dilabeli tidak berisiko. Setelah itu, sistem akan menghitung probabilitas prior, yaitu probabilitas awal untuk setiap kelompok. Kemudian, sistem menghitung probabilitas *likelihood* untuk setiap fitur (seperti usia, IMT, dan tekanan darah) menggunakan fungsi khusus yang menyertakan *Laplace Smoothing* untuk mencegah hasil probabilitas nol. Terakhir, sistem menghitung probabilitas posterior dengan mengalikan probabilitas prior dengan semua probabilitas *likelihood*. Hasilnya menunjukkan seberapa besar kemungkinan data pengguna termasuk ke dalam kelas "Berisiko" atau "Tidak Berisiko".

```

public function cetak()
{
    // Ambil hasil terbaru
    $latest = Deteksi::latest()->first();

    if (!$latest) {
        return redirect()->route('hasil')->with('error', 'Data hasil deteksi tidak ditemukan.');
    }

    // Load view ke PDF
    $pdf = Pdf::loadView('pages.cetak_hasil', compact('latest'));

    // Tampilkan di browser (preview untuk print)
    return $pdf->stream('hasil_deteksi_' . $latest->nama . '.pdf');
}

public function download()
{
    $latest = Deteksi::latest()->first();

    if (!$latest) {
        return redirect()->route('hasil')->with('error', 'Data hasil deteksi tidak ditemukan.');
    }

    $pdf = Pdf::loadView('pages.cetak_hasil', compact('latest'));
    return $pdf->download('hasil_deteksi_' . $latest->nama . '.pdf');
}

```

Gambar 4. 23 Proses cetak

Hasil prediksi akhir ditentukan dengan membandingkan kedua nilai probabilitas posterior. Kelas dengan nilai tertinggi menjadi hasil diagnosis. Hasil ini, bersama dengan data input asli dan nilai yang telah di-encode, kemudian disimpan ke dalam basis data melalui perintah Deteksi::create(). Setelah penyimpanan, pengguna diarahkan ke halaman hasil yang menampilkan status risiko mereka. Selain itu, fungsi tambahan seperti cetak() dan download() memungkinkan pengguna untuk membuat laporan dalam format PDF, yang sangat berguna untuk keperluan medis atau arsip pribadi.

4.4.3 Pengujian Sistem Menggunakan *Blackbox Testing*

Pengujian sistem merupakan tahap krusial untuk memastikan bahwa sistem yang telah dibangun berfungsi sesuai dengan kebutuhan dan spesifikasi yang telah ditentukan. Dalam penelitian ini, metode pengujian yang digunakan adalah *Blackbox Testing* atau pengujian kotak hitam. Metode ini berfokus pada fungsionalitas sistem tanpa melihat struktur internal atau kode program. Pengujinya hanya berinteraksi dengan antarmuka sistem (*user interface*) dan memastikan bahwa setiap input menghasilkan output yang diharapkan.

Fungsional	Skenario Uji	Hasil yang Diharapkan	Hasil
<i>Login</i> Pengguna	Memasukkan <i>email</i> dan <i>password</i> yang benar	Tampil halaman beranda dengan tambahan menu deteksi dan riwayat	Berhasil

	Memasukkan <i>email</i> dan <i>password</i> yang salah	Tampil error	keterangan login tidak berhasil	Berhasil
Deteksi	1. Memilih menu deteksi 2. Mengisi <i>form</i> identitas 3. Mengisi <i>form</i> deteksi 4. Submit <i>form</i> deteksi	1. Tampil halaman <i>form</i> deteksi 2. <i>Form</i> dapat terisi 3. Tampil <i>pop up</i> hasil deteksi	Berhasil	Berhasil
Riwayat	Memilih menu riwayat	Tampil halaman riwayat	Berhasil	Berhasil
<i>Logout</i> pengguna	Klik ikon di pojok kanan atas, klik <i>logout</i>	Tampil halaman depan	Berhasil	Berhasil

Hasil dari pengujian sistem menggunakan *Blackbox Testing* menunjukkan bahwa seluruh fungsionalitas utama sistem telah berjalan dengan baik sesuai dengan yang diharapkan. Pada fungsionalitas login pengguna, sistem berhasil memvalidasi kredensial pengguna, baik untuk kasus data yang benar maupun yang salah, sehingga pengguna dapat masuk atau menerima pesan kesalahan yang sesuai. Pada fungsionalitas deteksi, alur pengisian formulir hingga tampilan *pop-up* hasil deteksi juga berjalan dengan lancar. Pengguna dapat memilih menu deteksi, mengisi semua *field* yang diperlukan, dan menerima hasil deteksi yang akurat. Begitu pula dengan fungsionalitas riwayat dan *logout*, di mana sistem berhasil mengarahkan pengguna ke halaman yang sesuai.

Secara keseluruhan, hasil pengujian ini menegaskan bahwa sistem deteksi dini diabetes telah siap untuk digunakan karena semua fitur utamanya berfungsi dengan optimal dan memberikan output yang diharapkan dari setiap input yang diberikan oleh pengguna.

4.5 Menempatkan Sistem Deteksi Dini Penyakit *Diabetes Mellitus* pada *Server Hosting (Deployment)*.

Penerapan *Continuous Integration/Continuous Delivery* (CI/CD) bertujuan untuk mengotomatisasi penggabungan kode, pengujian, dan persiapan sistem untuk rilis, sehingga meminimalkan risiko kesalahan manusia dan mempercepat siklus *feedback* dari pengembangan ke lingkungan *staging* atau *production*. Dalam penelitian ini, penggunaan repositori GitHub menjadi fondasi utama. Praktik *Continuous Integration* (CI) dimulai ketika setiap

perubahan kode sumber di-push ke GitHub, yang seharusnya secara otomatis memicu *build* dan serangkaian test untuk memastikan fungsionalitas dan integrasi kode tetap terjaga. Setelah lolos pengujian, proses berlanjut ke *Continuous Delivery* (CD), di mana artifact yang siap di-deploy secara otomatis disiapkan. Meskipun paragraf berikut menjelaskan tahapan *deployment* yang terperinci, penting untuk ditekankan bahwa fondasi CI/CD inilah yang memastikan bahwa kode yang di-deploy adalah versi yang paling stabil, teruji, dan terkini, siap untuk diunggah ke *server hosting*.

Tahapan *deployment* dilaksanakan dengan mengunggah dan mengonfigurasi seluruh komponen sistem pada layanan server hosting yang memiliki koneksi stabil ke jaringan internet, sehingga sistem ini dapat diakses secara online oleh pengguna kapan pun dan di mana pun. Sebagai alamat akses utama, penelitian ini menggunakan domain utama mikpolije.com dengan subdomain sisdini.mikpolije.com sebagai alamat resmi sistem. Pemilihan struktur Uniform Resource Locator (URL) ini dirancang untuk mempermudah pengguna dalam mengingat dan mengakses layanan, sekaligus menegaskan identitas khusus dari sistem deteksi dini DM berbasis Naïve Bayes ini. Proses penempatan sistem ini mencakup beberapa langkah teknis, yaitu pengunggahan seluruh kode sumber program sistem ke *repository* GitHub, konfigurasi lingkungan *server* yang sesuai dengan kebutuhan sistem (termasuk *web server* dan runtime bahasa pemrograman), migrasi database yang berisi model *Naïve Bayes* yang sudah dilatih, dan terakhir adalah penyesuaian *Domain Name System* (DNS) untuk memastikan subdomain yang dipilih mengarah tepat ke lokasi sistem yang telah ter-deploy. Dengan berhasilnya tahap ini, Sistem Deteksi Dini Penyakit Diabetes Melitus dapat diakses secara online melalui alamat <https://sisdini.mikpolije.com> dan siap untuk digunakan oleh masyarakat luas.

BAB 5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan dari penelitian dengan judul Pernangan dan Implementasi Sistem Deteksi Dini Penyakit *Diabetes Mellitus* menggunakan Metode *Naïve Bayes* berbasis *Website* dapat ditarik kesimpulan:

1. Data rekam medis sebanyak 600 pasien, yang terdiri dari 300 pasien dengan diagnosis *Diabetes Mellitus* (DM) dan 300 pasien dengan diagnosis Hipertensi (HT), berhasil dikumpulkan dari Rumah Sakit Citra Husada. Delapan variabel utama yang relevan, meliputi usia, jenis kelamin, indeks massa tubuh (IMT), tekanan darah, kadar glukosa darah acak (GDA), riwayat hipertensi, riwayat kardiovaskular, dan aktivitas fisik, dipilih sebagai fitur klasifikasi. Proses *pre-processing* data dilakukan melalui tahapan *data cleaning*, *data selection*, dan *data transformation* sehingga menghasilkan data yang bersih, terstruktur, serta siap digunakan dalam pemodelan algoritma *Naïve Bayes*.
2. Perencanaan sistem deteksi dini *Diabetes Mellitus* dilakukan dengan pendekatan *Waterfall* model yang divisualisasikan melalui *Gantt Chart*. Perencanaan ini memberikan alur kerja yang sistematis mulai dari pengumpulan data, *pre-processing*, pemodelan, implementasi, hingga *deployment*. Dengan adanya rencana kerja tersebut, setiap tahapan penelitian dan pengembangan dapat dilaksanakan secara terstruktur, terukur, serta tepat waktu, sehingga meminimalisasi risiko keterlambatan maupun kesalahan dalam pembangunan sistem.
3. Pengujian algoritma *Naïve Bayes* menunjukkan *accuracy* terbaik sebesar 73,81% pada komposisi data latih dan data uji 65:35. Selain itu, hasil evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy* 77,5%, *precision* 100%, dan *recall* 72,7%, yang menunjukkan kemampuan model dalam mengklasifikasikan risiko diabetes secara cukup baik. Analisis dan desain sistem yang dikembangkan meliputi *flowchart*, *use case diagram*, *activity diagram*, *data flow diagram* (DFD), *entity relationship diagram*

(ERD), serta rancangan antarmuka. Hal ini memastikan sistem memiliki alur kerja yang jelas, basis data yang terintegrasi, serta tampilan yang mudah digunakan oleh pengguna maupun tenaga medis.

4. Sistem deteksi dini *Diabetes Mellitus* berhasil diimplementasikan dengan antarmuka yang sederhana, interaktif, dan sesuai prinsip *user experience* (UX) dan *user interface* (UI). Proses pengkodean memungkinkan validasi, normalisasi, dan transformasi data sehingga dapat diproses menggunakan algoritma *Naïve Bayes* untuk menghasilkan prediksi risiko secara akurat. Hasil *blackbox testing* menunjukkan bahwa seluruh fungsionalitas utama sistem, yaitu *login*, input faktor risiko, deteksi, riwayat deteksi, dan logout, berjalan optimal sesuai spesifikasi yang ditetapkan. Dengan demikian, sistem yang dikembangkan layak untuk digunakan sebagai alat bantu deteksi dini risiko *Diabetes Mellitus*.
5. Pelaksanaan *deployment* Sistem Deteksi Dini Penyakit *Diabetes Mellitus* berbasis *Website* menggunakan Metode *Naïve Bayes* berhasil dilakukan dengan mengunggah dan mengkonfigurasi seluruh komponen sistem ke *server hosting*, memastikan sistem dapat diakses secara online melalui alamat <https://sisdini.mikpolije.com> dan siap digunakan oleh pengguna.

5.2 Saran

Saran yang dapat diambil berdasarkan hasil dan pembahasan penlitian yang berjudul Perancangan dan Implementasi Sistem Deteksi Dini Penyakit *Diabetes Mellitus* Menggunakan Metode *Naïve Bayes* Berbasis *Website* yaitu:

1. Sistem deteksi dini diabetes ini dapat dikembangkan lebih kompleks lagi pada pemilihan variabel yang digunakan dengan menambahkan variabel faktor risiko lain yang relevan, sehingga item pertanyaan kepada pengguna bisa menjadi lebih detail dan komprehensif.
2. Peneliti selanjutnya disarankan dapat mengembangkan sistem deteksi dini diabetes ini dengan metode *machine learning* lain, seperti C4.5 atau *Support Vector Machine* (SVM), untuk menemukan aturan-aturan lain dan membandingkan hasil akurasinya. Hal ini juga perlu memperhatikan komposisi dataset yang digunakan, termasuk urutan data tersebut.

3. Sistem deteksi dini ini dapat dikembangkan dengan mengintegrasikannya kepada dinas terkait atau fasilitas kesehatan. Dengan demikian, apabila terdapat pasien yang terindikasi berisiko, dapat dilakukan *tracing* dan penanganan lebih lanjut secara langsung.
4. Sistem deteksi dini diabetes ini masih merupakan hasil klasifikasi berdasarkan data yang ada, sehingga apabila pasien memiliki kecurigaan gejala diabetes, sangat dianjurkan untuk segera melakukan pemeriksaan lebih lanjut dan berkonsultasi dengan tenaga medis profesional.

DAFTAR PUSTAKA

- Abdussomad, A., Kurniawan, I., & Wibowo, A. (2024). Prediksi Kemungkinan Penyakit Liver Menggunakan Algoritma Klasifikasi Naïve Bayes. *Technologia: Jurnal Ilmiah*, 15(3), Article 3. <Https://Doi.Org/10.31602/Tji.V15i3.15288>
- Ada. (2010). Diagnosis And Classification Of Diabetes Mellitus. *Diabetes Care*, 33(Suppl. 1). <Https://Doi.Org/10.2337/Dc10-S062>
- Ada. (2017). Standards Of Medical Care In Diabetes—2017 Abridged For Primary Care Providers. *Clinical Diabetes*, 35(1), 5–26. <Https://Doi.Org/10.2337/Cd16-0067>
- Afif, A. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus Di Rumah Sakit Aisyah. *Jurnal Ilmu Komputer Dan Matematika*, 1, 40–46.
- Al’ariq, Q., Permadi, G. S., Mashuri, C., & Andriani, A. (2024). Implementasi Naïve Bayes Dalam Memprediksi Penyakit Diabetes Mellitus. *Inovate: Jurnal Ilmiah Inovasi Teknologi Informasi*, 144–153.
- Ali. (2019). Rekayasa Perangkat Lunak. Cv Mfa.
- Ali Ikhwan, & Aslami, N. (2024). Implementasi Data Mining Untuk Manajemen Bantuan Sosial Menggunakan Algoritma K-Means. <Https://Core.Ac.Uk/Reader/478590070>
- Amran, R., Apriyani, A., & Dewi, N. P. (2022). Peran Penting Kelengkapan Rekam Medik Di Rumah Sakit. *Baiturrahmah Medical Journal*, 1(September 2021), 69–76.
- Arhami, M., & Nasir, M. (2020). Data Mining—Algoritma Dan Implementasi. Penerbit Andi.
- Bichri, H., Chergui, A., & Hain, M. (2024). Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets. *International Journal of Advanced Computer Science and Applications*, 15(2), 331–339. <https://doi.org/10.14569/IJACSA.2024.0150235>
- Bilous, R., & Donelly, R. (2014). Buku Pegangan Diabetes (Edisi Ke-4). Bumi Medika.
- Depkes RI. (2008). *Pedoman Teknis Penemuan Dan Tatalaksana DM* (pp. 1–43).

- Duri, V. R. S., Dian Ellina, A., & Nurwijayanti, N. (2024). Efektifitas Kepatuhan Pelaksanaan Empat Pilar Diabetes Militus Terhadap Kadar Gula Darah Dan Kadar Hba1c Penderita Diabetes Militus Tipe 2 (Studi Di Klinik Zahra Medika Silo Jember). *Klabat Journal Of Nursing*, 6(1), 14. <Https://Doi.Org/10.37771/Kjn.V6i1.990>
- Fadillah, M. A., Mulyani, E. D. S., & Ruuhwan, R. (2024). Komparasi Algoritma C4.5, Naive Bayes, K-Nearest Neighbor, Random Forest Untuk Prediksi Faktor Penyebab Penyakit Diabetes. *Indonesian Journal Of Digital Business*, 4(1), 37–46. <Https://Doi.Org/10.17509/Ijdb.V4i1.68999>
- Fatmawati, F., Wicaksono, S. A., & Wijoyo, S. H. (2023). Implementasi Algoritma Naïve Bayes Untuk Mendeteksi Resiko Tinggi Diabetes Melitus Pada Ibu Hamil (Studi Kasus: Puskesmas Kabupaten Malang). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(4), 851–856. <Https://Doi.Org/10.25126/Jtiik.20241046422>
- Fauzi, J. R. (2020). Algoritma Dan Flowchart Dalam Menyelesaikan Suatu Masalah.
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Twitter. *Smatika Jurnal : Stiki Informatika Jurnal*, 10(02), Article 02. <Https://Doi.Org/10.32664/Smatika.V10i02.455>
- Firdaus. (2021). Rancangan Alat Penyeduh Mie Otomatis Berbasis Arduino (Doctoral Dissertation, Politeknik Harapan Bersama Tegal).
- Freeman, A. M., Acevedo, L. A., & Pennings., N. (2023). National Library of Medicine. Diambil kembali dari Insulin Resistance: https://www.ncbi.nlm.nih.gov/sites/books/NBK507839/?utm_source
- Huda, M. (2019). Algoritma Data Mining: Analisis Data Dengan Komputer. Bisakimia.
- ID, I. D. (2021). Machine Learning : Teori, Studi Kasus dan Implementasi Menggunakan Python. *UNRI Press*. <https://doi.org/10.5281/zenodo.5113507>
- Idf. (2021). Diabetes.
- IDF. (2025a). Diabetes Facts and Figures. IDF. https://idf.org/about-diabetes/diabetes-facts-figures/?utm_source
- IDF. (2025b). IDF Diabetes Atlas. https://diabetesatlas.org/?utm_source
- Ihsan, M., & Adisasmita, A. C. (2023). Hubungan Hipertensi dengan Kejadian Diabetes Melitus Pada Populasi Perokok (Analisa Data IFLS 5) The

- Association Between Hypertension and Diabetes Melitus Among Smokers (Analysis Data of The Fifth Indonesia Family Live Survey). *Jurnal Dunia Kesmas*, 12(2), 51–61. <http://ejurnalmalahayati.ac.id/index.php/duniakesmas/index>
- Isnaini, N., & Ratnasari, R. (2018). Faktor risiko mempengaruhi kejadian Diabetes mellitus tipe dua. *Jurnal Kebidanan Dan Keperawatan Aisyiyah*, 14(1), 59–68. <https://doi.org/10.31101/jkk.550>
- Juventauricula, P., Hanggara, B. T., & Pramono, D. (2024). Pengembangan Sistem Informasi Point Of Sale (Pos) Berbasis Web Menggunakan Pendekatan Metode Waterfall (Studi Kasus: Restoran Altari). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 8(1), 97–106.
- Karyadiputra, E., & Setiawan, A. (2023). Sistem Pendukung Keputusan Berbasis Decision Tree Algorithm Untuk Prediksi Penyakit Diabetes. *Teknosains: Media Informasi Dan Teknologi*, 17(3), 294–301.
- Katmawanti, S., Ulfa, N. H., & Alma, L. R. (2022). Penyakit Diabetes Mellitus (L. R. Alma (ed.)). CV. Literasi Nusantara Abadi.
- Kemenkes RI. (2018). Laporan Nasional Riskesdas 2018. Jakarta: Badan Penelitian dan Pengembangan Kesehatan.
- Lanywati, Dr Endang. (2001). Diabetes Mellitus. Penerbit Kanisius.
- Listyawati, F. M. (2024). Perancangan Dan Pembuatan Sistem Deteksi Dini Penyakit Diabetes Melitus Di Puskesmas Sumbersari Jember Berbasis Website. Politeknik Negeri Jember.
- Mandalika, E. A., Dasuki, M., & Umilasari, D. R. (2024). *Implementasi Algoritma Naïve Bayes Menggunakan Pemilihan Atribut Information Gain Pada Penyakit Diabetes*. 5(4), 430–438.
- Maulana, M. (2009). Mengenal Diabetes: Panduan Praktis Menangani Penyakit Kencing Manis. Kata Hati.
- Muslim, Aziz, M., Prasetyo, B., & Marwani, E. L. H. (2019). Data Mining Algoritma C4.5. Semarang: Harian Jateng Network. <Http://Lib.Unnes.Ac.Id/Id/Eprint/33080> (June 2, 2023).
- Ningrum, F. C., Suherman, D., Aryanti, S., Prasetya, H. A., & Saifudin, A. (2019). Pengujian Black Box Pada Aplikasi Sistem Seleksi Sales Terbaik Menggunakan Teknik Equivalence Partitions. *Jurnal Informatika Universitas Pamulang*, 4(4), 125. <Https://Doi.Org/10.32493/Informatika.V4i4.3782>

- Nugroho, H., Ernawilis, Suheti, & Fadwa Syamlan, S. (2023). Penyuluhan Kesehatantentang Pengetahuan Pencegahan Diabetes Melitus Di Desa Rawat Rengas. *Jurnal Peduli Masyarakat*, 5(4), 1064.
- Nurrahmani, U. (2012). Stop! Diabetes. Familia.
- Nurrahmani, U. (2015). Stop! Hipertensi. Familia.
- Nurussakinah, N. (2024). Implementasi Algortima Random Forest Dan Teknik Smote Untuk Klasifikasi Penyakit Diabetes [Undergraduate, Universitas Islam Negeri Maulana Malik Ibrahim]. <Http://Etheses.Uin-Malang.Ac.Id/65251/>
- P2PTM. (2008). Pedoman Pengendalian Diabetes Melitus Dan Penyakit Metabolik. In *Indonesia Sehat 2010* (p. 82). file:///D:/ebook/dsa664.pdf
- Paleva, R. (2019). Mekanisme Resistensi Insulin Terkait Obesitas. *Jurnal Ilmiah Kesehatan Sandi Husada*, 357.
- Panna, Firda, Mahakena, I. H., Pratiw, N., Rambu, A., & Agustina, V. (2021). “Deteksi Risiko Diabetes Mellitus Tipe 2 Menggunakan Findrisc.” *Magistrorum Et Scholarium: Jurnal Pengabdian Masyarakat* 22: 357–62.
- Prabowo, M. (2020). Metodologi Pengembangan Sistem Informasi. Lp2m Press Iain Salatiga.
- Prasetyo, M. Z., Susanto, E. R., & Wantoro, A. (2023). Sistem Informasi Rekam Medis Pasien Thalassemia (Studi Kasus : Popti Cabang Bandar Lampung). 4(3), 349–355.
- Pressman, Roger, & Maxim. (2015). Software Engineering. 8th Ed. New York: McGraw-Hill.
- Pressman, R. S. (2010). *Software Engineering: A Practitioner’s Approach*. McGraw-Hill Education. <https://doi.org/10.1145/336512.336521>
- Putra, I. W. G. >Artawan E., Kurniasari, N. M. D., Dewi, N. P. E. P., Suarjana, I. K., Duana, I. M. K., Mulyawan, I. K. H., Riono, P., Alisjahbana, B., Probandari, A., Notobroto, H. B., & Wahyuni, C. U. (2019). The Implementation Of Early Detection In Tuberculosis Contact Investigation To Improve Case Finding. *Journal Of Epidemiology And Global Health*, 9(3), 191–197. <Https://Doi.Org/10.2991/Jegh.K.190808.001>
- Rahmansyah, N. (2024). Analisa Prediksi Penyakit Diabetes Menggunakan Metode Naive Bayes Dan K-Nn. 5.

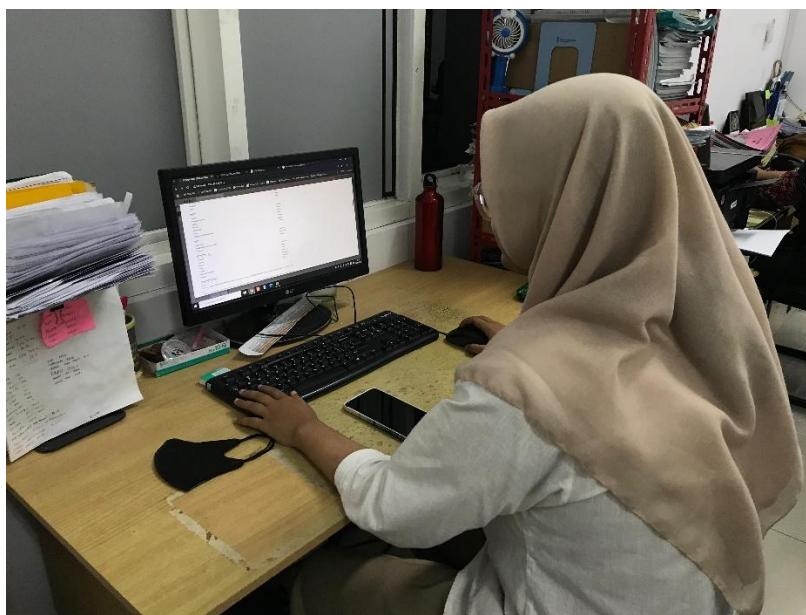
- Riskesdas. (2018). Laporan Riskesdas 2018 Nasional.Pdf. In Lembaga Penerbit Balitbangkes (P. Hal 156).
- Romadloni, P. L., Kusuma, B. A., & Baihaqi, W. M. (2022). Komparasi Metode Pembelajaran Mesin Untuk Implementasi Pengambilan Keputusan Dalam Menentukan Promosi Jabatan Karyawan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 623–628.
- Rozi, M. F. (2023). Penerapan Data Mining Menggunakan Metode Naive Bayes Untuk Klasifikasi Data Penentuan Hasil Penjualan Dalam Strategi Pemasaran. *Jurnal Komputer Teknologi Informasi Dan Sistem Informasi (Juktisi)*, 2(2), Article 2. <Https://Doi.Org/10.62712/Juktisi.V2i2.137>
- Salsabila, L. N., Dwi Pangga, M. R., Yasser, S. M., Riyani, N. A., Aminah, S., & Wahyunengsih, W. (2024). Application of Naïve Bayes Algorithm for Diabetes Prediction. *Unisda Journal of Mathematics and Computer Science (UJMC)*, 10(1), 56–68. <https://doi.org/10.52166/ujmc.v10i1.6886>
- Sari, & Abdilah. (2019). Web Programming.
- Sulistyowati, E. (2016). Ilmu Gizi: Teori & Aplikasi (Hardinsyah & I. D. N. Supariasa, Eds.). Penerbit Buku Kedokteran Egc.
- Tamornpark, R., Utsaha, S., Apidechkul, T., Panklang, D., Yeemard, F., & Srichan, P. (2022). Quality Of Life And Factors Associated With A Good Quality Of Life Among Diabetes Mellitus Patients In Northern Thailand. *Health And Quality Of Life Outcomes*, 20(1), 1–11. <Https://Doi.Org/10.1186/S12955-022-01986-Y>
- Triandita, N., R. Zakaria, F., Prangdimurti, E., & Eska Putri, N. (2016). Perbaikan Status Antioksidan Penderita Diabetes Tipe 2 Dengan Tahu Kedelai Hitam Kaya Serat. *Jurnal Teknologi Dan Industri Pangan*, 27(2), 123–130. <Https://Doi.Org/10.6066/Jtip.2016.27.2.123>
- Usman, M. S., Khan, M. S., & Butler, J. (2021). Chronic Kidney Disease and Type 2 Diabetes. American Diabetes Association. <https://www.ncbi.nlm.nih.gov/books/NBK571718/>
- Wahid, & Abdul., A. (2020). “Analisis Metode Waterfall Untuk Pengembangan Sistem Informasi.” *Jurnal Lmu-Ilmu Informatika Dan Manajemen Stmik*: 1– 5.
- Who. (2024). Diabetes. Who.

LAMPIRAN

Lampiran 1 Lembar Observasi

	<p style="text-align: center;">LEMBAR OBSERVASI</p> <p style="text-align: center;">PERANCANGAN DAN IMPLEMENTASI SISTEM DETEKSI DINI</p> <p style="text-align: center;">PENYAKIT DIABETES MELLITUS MENGGUNAKAN METODE NAÏVE</p> <p style="text-align: center;">BAYES BERBASIS DATA REKAM MEDIS</p>												
<p>Petunjuk pengisian:</p> <ol style="list-style-type: none"> 1. Beri tanda “v” pada lembar observasi apabila rekam medis yang diamati memuat informasi faktor risiko berikut. 2. Tulis kolom diagnosa dengan “DM” apabila diagnose dari dokter adalah <i>Diabetes Mellitus</i> (DM). Selain diagnosis Diabetes Mellitus (DM) tulis kolom diagnosa dengan “Non DM” 													
No	No RM	Usia	IMT	JK	Riwayat keluarga	Aktifitas fisik	Hipertensi	Dislipidemia	Riwayat terdahulu	Diet	GDA	Merokok	

Lampiran 2 Dokumentasi Studi Pendahuluan



Lampiran 3 Surat Rekomendasi Bankesbangpol



PEMERINTAH KABUPATEN JEMBER
BADAN KESATUAN BANGSA DAN POLITIK

Jalan Letjen S Parman No. 89 Telp. 337853 Jember

Kepada
 Yth. Sdr. Direktur Rumah Sakit Citra Husada
 Jember.
 di -
 Jember

SURAT REKOMENDASI

Nomor : 074/3292/415/2023

Tentang

PENELITIAN

Dasar : 1. Permendagri RI Nomor 7 Tahun 2014 tentang Perubahan atas Permendagri RI Nomor 64 Tahun 2011 tentang Pedoman Penerbitan Rekomendasi Penelitian
 2. Peraturan Bupati Jember No. 46 Tahun 2014 tentang Pedoman Penerbitan Surat Rekomendasi Penelitian Kabupaten Jember

Memperhatikan : Surat Politeknik Negeri Jember, 16 November 2023, Nomor: 18112, Perihal: Permohonan Rekomendasi

MEREKOMENDASIKAN

Nama : Rada Aulia Irawan
 NIM : G4121123
 Daftar Tim : Daftar terlampir
 Instansi : Manajemen Informasi Kesehatan
 Alamat : Jl. Maspipi, Krajan Timur, Sumbersari, Kec. Sumbersari, Kabupaten Jember, Jawa Timur 68121
 Keperluan : Melaksanakan kegiatan penelitian dengan jufu/terkait Penelitian Tugas Akhir
 Lokasi : Rumah Sakit Citra Husada Jember
 Waktu Kegiatan : 16 November 2023 s/d 31 Desember 2025

Apabila tidak bertentangan dengan kewenangan dan ketentuan yang berlaku, diharapkan Saudara memberi bantuan tempat dan atau data seperlunya untuk kegiatan dimaksud.
 1. Kegiatan dimaksud benar-benar untuk kepentingan Pendidikan.
 2. Tidak dibenarkan melakukan aktivitas politik.
 3. Apabila situasi dan kondisi wilayah tidak memungkinkan akan dilakukan penghentian kegiatan.
 Demikian atas perhatian dan kerjasamanya disampaikan terimakasih.

Dititapkan di : Jember
 Tanggal : 16 November 2023
KEPALA BAKESBANG DAN POLITIK
KABUPATEN JEMBER

Ditandatangani secara elektronik



Drs. SIGIT AKBARI, M.Si.
PEMBINA UTAMA MUDA
NIP. 19650309 198602 1 002

Tembusan :
 Yth. Sdr. 1. Direktur Politeknik Negeri Jember.
 2. Yang Bersangkutan.

Lampiran 1

Surat Nomor : 074/3292/415/2023
 Tanggal : 16 November 2023
 Daftar Tim : 1. Putri Nur Hidayah
 2. Ana Tazkia Zahro
 3. Fasya Nur Alvia Kusnia
 4. Khana Amita Sofir
 5. Rada Aulia Irawan
 6. Kavita Wulanandika
 7. Amellinda Anisa Marsel
 8. Safira Damayanti
 9. Flora Puspita Nirmala Sari
 10. Sekar Ayu Surya Kinanti
 11. Fattika Fizatullah Mustafa
 12. Brenda Rizky Tsania Gunawan

Lampiran 4 Surat Ijin Penelitian



**KEMENTERIAN PENDIDIKAN TINGGI, SAINS,
DAN TEKNOLOGI
POLITEKNIK NEGERI JEMBER**

Jalan Mastrip Kotak Pos 164 Jember 68121 Telp. (0331) 333532-34
Email : politeknik@polije.ac.id Website : <https://www.polije.ac.id>

Nomor : **13306** /PL17/PP/2025
 Lampiran : -
 Perihal : Ijin Penelitian

06 AUG 2025

**Yth. Direktur RS. Citra Husada
 Jl. Teratai No.22, Gebang Timur
 Jember**

Sehubungan dengan pelaksanaan Tugas Akhir pada kegiatan akademik Politeknik Negeri Jember Tahun Akademik 2024/2025 Program Studi Sarjana Terapan Manajemen Informasi Kesehatan Jurusan Kesehatan, dengan hormat kami mohon agar mahasiswa kami diperkenankan melaksanakan penelitian di RS. Citra Husada Jember yang Bapak/Ibu pimpin. Adapun nama mahasiswa dan judul penelitian :

NO	NIM	NAMA	JUDUL
1.	G41210047	Brenda Rizky Tsania Gunawan (0858-1233-6525)	Perancangan dan Implementasi Sistem Deteksi Dini Penyakit Diabetes Mellitus Menggunakan Metode <i>Naïve Bayes</i> Berbasis Website

Contact person Koordinator Tugas Akhir yang dapat dihubungi dengan Ibu Indah Muflihatin, S.Si.T., M.Kes (0821-3249-3080).

Demikian atas perhatian dan kerjasamanya disampaikan terima kasih.



Lampiran 5 Surat permohonan rekomendasi



**KEMENTERIAN PENDIDIKAN TINGGI, SAINS,
DAN TEKNOLOGI
POLITEKNIK NEGERI JEMBER**

Jalan Mastrip Kotak Pos 164 Jember 68121 Telp. (0331) 333532-34
Email : politeknik@polije.ac.id Website : <https://www.polije.ac.id>

Nomor : **13305** /PL17/PP/2025
Lampiran : -
Perihal : Permohonan Rekomendasi

06 AUG 2025

**Yth. Kepala Badan Kesatuan Bangsa dan Politik
Pemerintah Kabupaten Jember
Jln. S. Parman No 89
Jember**

Sehubungan dengan pelaksanaan Tugas Akhir pada kegiatan akademik Politeknik Negeri Jember Tahun Akademik 2024/2025 Program Studi Sarjana Terapan Manajemen Informasi Kesehatan Jurusan Kesehatan, dengan hormat mohon ijin rekomendasi agar mahasiswa kami dapat melaksanakan penelitian di RS. Citra Husada Jember. Adapun nama mahasiswa dan judul penelitian:

NO	NIM	NAMA	JUDUL
1.	G41210047	Brenda Rizky Tsania Gunawan (0858-1233-6525)	Perancangan dan Implementasi Sistem Deteksi Dini Penyakit Diabetes Mellitus Menggunakan Metode Naïve Bayes Berbasis Website

Contact person Koordinator Tugas Akhir yang dapat dihubungi dengan Ibu Indah Muflihatin, S.Si.T., M.Kes (0821-3249-3080).

Demikian atas perhatian dan kerjasamanya disampaikan terima kasih.



Lampiran 6 Surat Keterangan Pengajuan Etik



**KEMENTERIAN PENDIDIKAN TINGGI, SAINS,
DAN TEKNOLOGI**
POLITEKNIK NEGERI JEMBER
PUSAT PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT
 Jalan Mastrip Kotak Pos 164 Jember – 68101 Telp. (0331) 333532-333534
 Fax. (0331) 333531 e-mail :Polije@ac.id laman : www.polije.ac.id

KETERANGAN PERSETUJUAN ETIK

ETHICAL APPROVAL

Nomor: 1097/PL17.4/PG/ 2025

Komisi Etik, Politeknik Negeri Jember dalam upaya melindungi hak asasi dan kesejahteraan subjek penelitian manusia, telah mengkaji dengan teliti protokol berjudul:

The Ethics Committe of the State Polytechnic Of Jember, With regards of the protection of human rights and welfare in medical reseach, has carefully reviewed the proposal entitled:

'PERANCANGAN DAN IMPLEMENTASI SISTEM DETEKSI DINI PENYAKIT DIABETES MELLITUS MENGGUNAKAN METODE NAÏVE BAYES BERBASIS WEBSITE.'

Nama Peneliti Utama

: Brenda Rizky Tsania Gunawan.: G41210047

Nama Institusi

: Program Studi Manajemen Informasi Kesehatan
Jurusan Kesehatan

Name of institution

Politeknik Negeri Jember

Dan telah menyetujui protokol tersebut diatas.

And approved the above mentioned proposal.

Jember, 1st Juni 2025

Menyetujui

Koordinator Komisi Etik
Bidang Kesehatan

Maya Weka Santi, S.KM, M.Kes
NIP. 19920515 201803 2 001

Mengetahui
Ketua Komisi Etik Penelitian
Bidang Hewan Coba dan Kesehatan

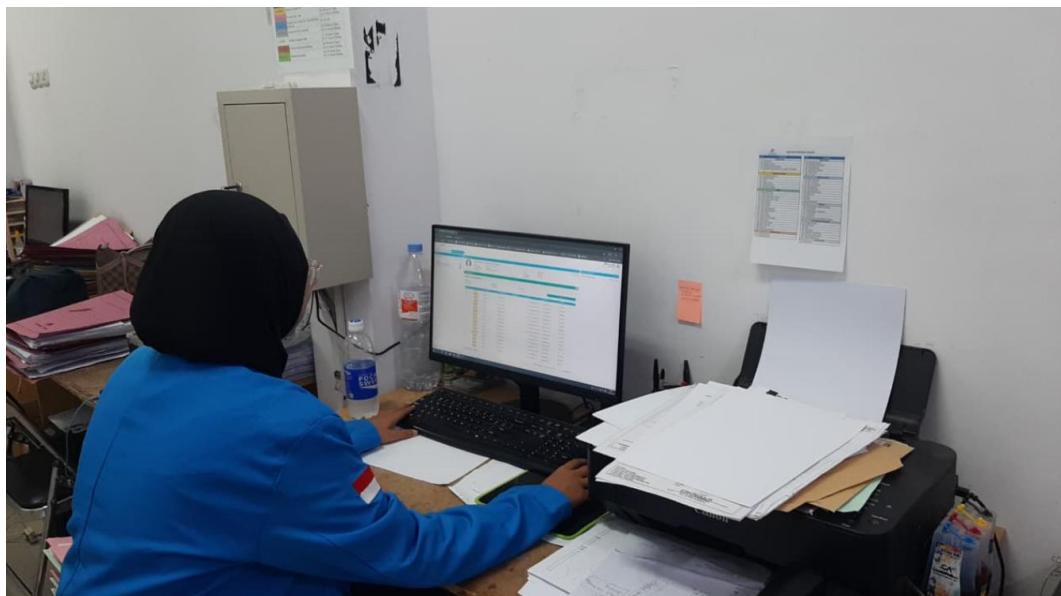


Dr. Ir. Hariadi Subagja,S.Pt,M.P, IPM
NIP. 19701213 199703 1 002

Lampiran 7 Sertifikat turnitin



Lampiran 8 Dokumentasi penelitian



Lampiran 9 Surat balasan dari sakit


RUMAH SAKIT CITRA HUSADA JEMBER
 Jl. Teratai No. 22 Jember
 Telp. (0331) 486200 Fax. (0331) 427088
 Website : www.rscitrahusada.com Email : rs_citrahusada@yahoo.co.id
★★★★★


Jember, 08 Agustus 2025

Nomor	:	1145/ RSCH/ VIII/ 2025
Sifat	:	Penting
Lampiran	:	-
Perihal	:	Pemberitahuan

Kepada Yth.

Direktur Politeknik Negeri Jember

Cq. Ketua Program Studi Manajemen Informasi Kesehatan

Di

Tempat

Menindak lanjuti surat saudara Nomor: 13306/PL17/PP/2025 tanggal 06 Agustus 2025 perihal Permohonan Izin Penelitian serta Surat dari Badan Kesatuan Bangsa dan Politik Nomor : 074/3292/415/2023 tanggal 16 November 2023 perihal Penelitian. Dengan ini kami menyetujui untuk mahasiswa saudara melakukan penelitian tersebut a.n. Brenda Rizky Tsania Gunawan NIM: G41210047 dengan Judul Penelitian "**Perancangan dan Implementasi Sistem Deteksi Dini Penyakit Diabetes Mellitus Menggunakan Metode Naïve Bayes Berbasis Website**". Dengan mengikuti segala peraturan yang telah ditentukan oleh Rumah Sakit Citra Husada Jember.

Demikian, atas perhatian dan kerjasama saudara kami sampaikan terima kasih.

Rumah Sakit Citra Husada Jember


**RUMAH SAKIT
CITRA HUSADA**
dr. Fatkhur Ruli Malik Qilsi
 Direktur

Tembusan, Yth :

- 1. Kabid. Penunjang Medik
- 2. Komite Etik Penelitian
- 3. Ka. Unit Rekam Medik
- 4. Mahasiswa Ybs