

Sistem Deteksi Dini *Diabetes Mellitus* Berdasarkan Rekam Medis Menggunakan Algoritma *K-Nearest Neighbor*

Arleni Aulia Yunitasari¹, Mudafiq Riyan Pratama², Muhammad Yunus³, Gandu Eko Julianto Suyoso⁴

¹Manajemen Informasi Kesehatan, Politeknik Negeri Jember, [auliaarleni24@gmail.com](mailto:auliarleni24@gmail.com)

²Manajemen Informasi Kesehatan, Politeknik Negeri Jember, mudafiq.riyan@polije.ac.id

³Manajemen Informasi Kesehatan, Politeknik Negeri Jember, m.yunus@polije.ac.id

⁴Manajemen Informasi Kesehatan, Politeknik Negeri Jember, gandu.eko.js@polije.ac.id

Keywords

Diabetes Mellitus,
Early Detection,
K-Nearest Neighbor,
Classification,
Medical Records

ABSTRACT

At RSD dr. Soebandi Jember, 51% of Diabetes Mellitus (DM) patients are diagnosed after complications occur, and DM is the third leading cause of death among non-communicable diseases, accounting for 13.6%. This situation indicates a high rate of delayed case identification. Delayed diagnosis significantly increases patient mortality and morbidity rates, emphasizing the urgent need for an effective, integrated early, and detection system. This study developed a web-based early detection system for DM using the K-Nearest Neighbor (K-NN) algorithm with the Waterfall development method, consisting of the stages of communication, planning, modeling, construction, and deployment. The data comprised from 342 inpatient medical records, and after preprocessing, 164 clean data were obtained with variables including age, gender, family history, blood pressure, random blood sugar, and body mass index. The data were split using stratified sampling (50:50), with K=5 value selected based on the best performance. Blackbox testing was conducted to ensure the system's functionality, while performance testing compared the system's classification results with the test data. The performance of the K-NN algorithm for DM detection was evaluated using a Confusion Matrix, resulting in an accuracy of 97.56%, precision of 100%, and recall of 95.83%, which were consistent with the results from the WEKA tool. This system is expected to serve as an early screening tool and support DM prevention efforts.

Kata Kunci

Diabetes Melitus,
Deteksi Dini,
K-Nearest Neighbor
Klasifikasi,
Rekam Medis

ABSTRAK

Di RSD dr. Soebandi Jember, sebanyak 51% pasien *Diabetes Mellitus* (DM) terdiagnosis setelah komplikasi, dan DM menjadi penyebab kematian ketiga tertinggi penyakit tidak menular dengan persentase 13,6%. Kondisi ini menunjukkan masih tingginya keterlambatan dalam proses identifikasi kasus. Diagnosis yang terlambat ini secara signifikan meningkatkan angka kematian dan kesakitan pasien, menekankan kebutuhan mendesak akan sistem deteksi dini yang efektif dan terintegrasi. Penelitian ini bertujuan mengembangkan sistem deteksi dini DM berbasis website menggunakan algoritma K-Nearest Neighbor (K-NN) dengan metode pengembangan *Waterfall*, meliputi tahapan *communication*, *planning*, *modeling*, *construction*, *deployment*. Data berasal dari 342 rekam medis pasien rawat inap, dan setelah preprocessing diperoleh 164 data bersih dengan variabel usia, jenis kelamin, riwayat keluarga, tekanan darah, gula darah acak, dan indeks massa tubuh. Data dibagi dengan stratified sampling (50:50), menggunakan nilai K=5 berdasarkan performa terbaik. Pengujian blackbox memastikan fungsi sistem berjalan baik, sedangkan pengujian kinerja membandingkan hasil klasifikasi sistem dengan data uji. Hasil pengujian kinerja algoritma K-NN untuk pendeteksian penyakit *Diabetes Mellitus* dievaluasi menggunakan Confusion Matrix, dan didapatkan nilai accuracy sebesar 97,56%, precision 100%, dan *recall* sebesar 95,83%, serta sesuai dengan hasil tools WEKA. Sistem ini diharapkan menjadi alat skrining awal dan mendukung pencegahan DM.

Korespondensi Penulis:

Arleni Aulia Yunitasari,
Politeknik Negeri Jember,
Jl. Mastrip PO BOX 164 Jember, Jawa Timur
Telepon : +6281334605948
Email: [auliaarleni24@gmail.com](mailto:auliarleni24@gmail.com)

Submitted : 03-01-2026; Accepted : 26-02-2026;
Published : 26-02-2026

Copyright (c) 2024 The Author (s) This article is distributed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

1. PENDAHULUAN

Diabetes Mellitus (DM) merupakan penyakit metabolik kronis yang ditandai dengan kadar gula darah lebih tinggi dari nilai normal [1]. DM merupakan salah satu dari empat Penyakit Tidak Menular (PTM) prioritas, dan penyebab utama kebutaan, serangan jantung, *stroke*, gagal ginjal dan amputasi anggota tubuh bagian bawah. DM menempati urutan ke sembilan dari sepuluh penyakit penyebab kematian teratas di dunia [2]. Data *International Diabetes Federation* (IDF) menemukan bahwa pada tahun 2021 jumlah kasus diabetes di Indonesia mengalami peningkatan signifikan selama sepuluh tahun terakhir, dengan lonjakan sebesar 167% dari 7,29 juta kasus pada tahun 2011 menjadi 19,47 juta kasus pada tahun 2021. Jumlah tersebut diproyeksikan meningkat menjadi 28,57 juta orang pada tahun 2045 atau naik sebesar 47% dibandingkan dengan jumlah 19,47 juta orang pada tahun 2021 [3]. Hasil Riset Kesehatan Dasar (Riskesdas) tahun 2018 menunjukkan bahwa prevalensi DM berdasarkan diagnosis dokter pada penduduk semua umur menurut provinsi tahun 2018, provinsi Jawa Timur menempati urutan kelima dengan prevalensi 2,6 % [4]. Berdasarkan profil Dinas Kesehatan Provinsi Jawa Timur tahun 2021, terdapat 929.810 kasus DM dan Kabupaten Jember masuk peringkat tertinggi ke-8 dengan jumlah kasus mencapai 37.786 dari total keseluruhan [5].

Kasus kematian akibat *Diabetes Mellitus* di Indonesia berada pada peringkat keenam setelah Jepang, Pakistan, India, Amerika Serikat, dan Tiongkok dengan angka kematian akibat diabetes di Indonesia tercatat sebanyak 236 ribu pada tahun 2021 [6]. DM merupakan penyakit dengan angka kematian yang cukup banyak dibandingkan penyakit tidak menular lainnya, dibuktikan dengan data persentase kematian di RSD dr. Soebandi tahun 2024 sebesar 13,6%, dikarenakan hal tersebut peneliti memilih penyakit DM. Di RSD dr. Soebandi Jember, sebanyak 51% pasien DM terdiagnosis setelah terjadi komplikasi. Padahal, jika didiagnosis sejak dini, pengobatan diabetes dapat segera dilakukan dan komplikasi penyakit berbahaya lainnya bisa dihindari [7]. Hal ini dapat didukung dengan pembuatan sistem deteksi dini DM berdasarkan rekam medis pasien menggunakan teknik *data mining* untuk menggali informasi dari kumpulan data *Diabetes Melitus*.

Data mining adalah metode analisis data melalui penggalian data untuk memperoleh data dengan cara yang berbeda dari sebelumnya [8]. *Supervised Learning* merupakan salah satu teknik *Data Mining*. Salah satu jenis dari *Supervised Learning* adalah klasifikasi. Klasifikasi digunakan untuk menentukan keputusan sesuai dengan pola baru yang didapat dari pola data lama menggunakan perhitungan algoritma. Salah satu algoritma klasifikasi yang bekerja dengan cara mengukur kedekatan antara data baru dan data lama berdasarkan bobot data yang sudah ada adalah algoritma *K-Nearest Neighbor* (K-NN) [7]. Metode K-NN memiliki sejumlah keunggulan, di antaranya pelatihan sangat cepat, sederhana, dan efektif pada data pelatihan yang besar [9]. Selain itu, K-NN mempunyai beberapa keunggulan yaitu tahan terhadap data latih yang *Noisy* dan efektif digunakan pada data latih berukuran besar [10].

Penelitian yang dilakukan oleh Kurniadi *et al.* pada tahun 2021 yang berjudul “Perbandingan Akurasi Algoritma *K-Nearest Neighbor* Dan *Logistic Regression* Untuk Klasifikasi Penyakit Diabetes” memperoleh nilai akurasi algoritma K-NN 85,06% dan algoritma *Logistic Regression* 77,92% [7]. Hasil ini mengindikasikan bahwa algoritma K-NN memiliki akurasi dan kinerja yang lebih unggul dibandingkan algoritma *Logistic Regression* dalam mengklasifikasikan penyakit diabetes. Pada penelitian Amien *et al.* pada tahun 2023 yang berjudul “Perbandingan Metode *Naïve Bayes* dan K-NN (*K-Nearest Neighbor*) dalam Klasifikasi Penyakit Diabetes” juga mendapatkan nilai akurasi tertinggi dari penggunaan metode K-NN dengan K=5 sebesar 90%, lalu nilai akurasi dari *Naïve Bayes* sebesar 80% [9]. Penelitian Loka & Marsal tahun 2023 juga memperoleh nilai akurasi dari algoritma K-NN sebesar 96,10 % sedangkan pada algoritma NBC sebesar 90,94%, sehingga pada penelitian ini, penulis memilih menggunakan algoritma K-NN [11].

Berdasarkan penelitian terdahulu, dapat diketahui bahwa penelitian yang akan dilakukan oleh peneliti memiliki persamaan metode dengan penelitian dari Dwi Fasnuari, dkk. (2022) yaitu menggunakan algoritma K-NN, dan perbedaannya terletak pada hasil penelitian peneliti berupa sistem deteksi dini untuk penyakit *Diabetes Mellitus*, sedangkan penelitian dari Dwi Fasnuari, dkk. (2022) hanya menganalisis akurasi metode K-NN untuk deteksi penyakit *Diabetes Mellitus*. Adapun perbedaan dengan penelitian dari Nuradha, dkk. (2023) yaitu pada metode yang digunakan dalam membuat klasifikasi *Diabetes Mellitus* yaitu menggunakan algoritma C4.5. Selain hal tersebut, terdapat pula perbedaan dari segi output penelitian yang menghasilkan hasil akurasi, sedangkan hasil dari penelitian ini berupa sistem deteksi dini. Selain perbedaan metode dan output, penelitian ini juga memiliki perbedaan pada variabel yang digunakan, yaitu penambahan variabel riwayat *Diabetes Mellitus* dalam keluarga, yang belum terdapat pada penelitian sebelumnya.

Berdasarkan latar belakang yang telah diuraikan, maka dilakukan penelitian yang berjudul “Penerapan Algoritma *K-Nearest Neighbor* Untuk Deteksi Dini *Diabetes Mellitus* Berdasarkan Rekam Medis Pasien Rawat Inap di RSD dr. Soebandi Jember”. Penelitian ini menggunakan suatu teknik dalam *data mining* yaitu klasifikasi *K-Nearest Neighbor* (K-NN) dengan menggunakan 6 variabel berdasarkan faktor risiko diabetes yaitu riwayat penyakit *Diabetes Mellitus* dalam keluarga, usia, jenis kelamin, kadar gula darah acak, tekanan darah, serta indeks massa tubuh (IMT). Sistem deteksi dini ini diharapkan mampu mencapai tingkat akurasi yang lebih baik dibandingkan penelitian sebelumnya, sehingga diharapkan dapat membantu mencegah dampak terburuk yang mungkin timbul akibat penyakit *Diabetes Mellitus*.

2. METODE PENELITIAN

2.1 Jenis Penelitian

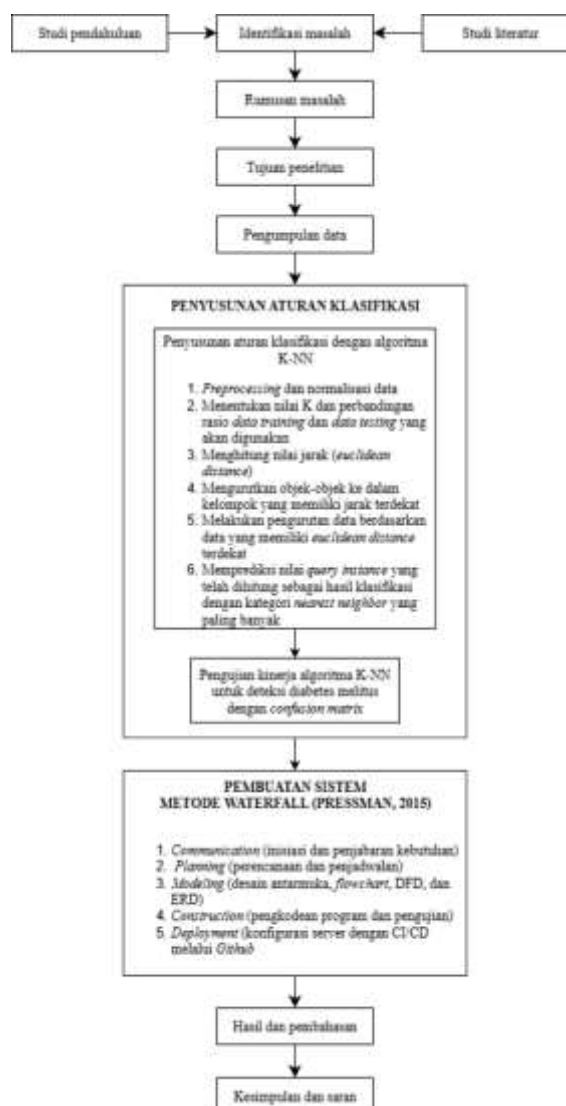
Jenis penelitian ini adalah penelitian kuantitatif yang diwujudkan dalam bentuk pengembangan sistem berbasis *Website* dengan menerapkan teknik klasifikasi menggunakan algoritma K-NN serta dilanjutkan dengan pengembangan sistem deteksi dini menggunakan metode *Waterfall*.

2.2 Teknik Pengumpulan Data

Pengumpulan data dilakukan dengan observasi data sekunder berupa rekam medis pasien rawat inap di Rumah Sakit Daerah (RSD) dr. Soebandi, Kabupaten Jember. Observasi dilakukan untuk mengumpulkan *dataset* yang akan digunakan dalam pembuatan aturan dengan algoritma K-NN. Jumlah data yang digunakan dalam penelitian ini adalah 164 *dataset*. Jumlah data minimal yang digunakan dalam teknik *data mining* adalah 100 data [12]. Hal ini sejalan dengan penelitian yang dilakukan oleh Lubis et al. (2023), Romlah & Fatah (2024), dan Dwi Fasnuri et al. (2022), yang menggunakan ≥ 100 data dan menghasilkan nilai akurasi yang tinggi [13][14][8]. Maka dari itu, dapat dikatakan bahwa pada teknik klasifikasi *data mining* minimal menggunakan 100 data.

Penelitian ini menggunakan data sekunder sebagai sumber data, yaitu berupa rekam medis rawat inap tahun 2021-2024 sebanyak 342 data kotor dengan diagnosa *Diabetes Mellitus* sebanyak 246 dan diagnosa hipertensi sebanyak 96. Rekam medis yang digunakan pada penelitian ini adalah rekam medis pasien rawat inap di Rumah Sakit Daerah (RSD) dr. Soebandi dengan diagnosa *Diabetes Mellitus* (kode ICD-10 E10.-, E11.-, E12.-, E13.-, dan E14.-) serta rekam medis lain dengan diagnosa hipertensi (kode ICD 10 I10). Adapun formulir yang digunakan sebagai sumber informasi antara lain resume medis pasien, assesment gawat darurat, pengkajian awal keperawatan rawat inap, asuhan gizi, catatan perkembangan pasien terintegrasi, dan hasil pemeriksaan laboratorium. Penelitian ini menggunakan suatu teknik dalam *data mining* yaitu klasifikasi *K-Nearest Neighbor* (K-NN) dengan menggunakan 6 variabel berdasarkan faktor risiko diabetes yaitu riwayat penyakit *Diabetes Mellitus* dalam keluarga, usia, jenis kelamin, kadar gula darah acak, tekanan darah, serta indeks massa tubuh (IMT).

2.3 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Berikut merupakan penjelasan dari tahapan penelitian yang dilakukan:

- Studi pendahuluan, yaitu kegiatan awal dalam pengamatan secara langsung mengenai kondisi dari tempat penelitian untuk mengetahui gambaran yang jelas dan pengumpulan data-data yang berkaitan.
- Studi literatur, yaitu mengumpulkan dan menggali informasi dari beberapa referensi yang berkaitan dengan permasalahan yang akan dibahas. Sumber literatur yang digunakan diperoleh dari penelitian terdahulu, buku, laporan, dan regulasi yang berkaitan dengan penelitian.
- Identifikasi masalah, yaitu tahapan penelitian untuk menentukan dan mengidentifikasi masalah yang berhubungan dengan kejadian Diabetes Melitus, khususnya untuk daerah Kabupaten Jember. Identifikasi masalah akan menghasilkan rumusan masalah, tujuan penelitian, dan manfaat penelitian.
- Rumusan masalah, yaitu merumuskan masalah terkait pembuatan sistem deteksi dini *Diabetes Mellitus* berdasarkan rekam medis pasien rawat inap berdasarkan metode yang akan digunakan.
- Tujuan penelitian, yaitu menentukan tujuan yang akan dicapai dalam proses penelitian serta untuk menentukan hasil penelitian dan mempertajam pembahasan.
- Pengumpulan data, yaitu mengumpulkan data yang diperoleh dengan observasi rekam medis pasien rawat inap di Rumah Sakit Daerah (RSD) dr. Soebandi, Kabupaten Jember. Data yang diperoleh kemudian digunakan sebagai *data training* dalam pembuatan aturan klasifikasi.
- Penyusunan aturan klasifikasi menggunakan algoritma K-NN, meliputi tahapan *preprocessing*, normalisasi data, klasifikasi dengan *tools* WEKA, dan pengujian kinerja dengan *Confusion Matrix*.
- Pembuatan sistem, dilakukan menggunakan metode *Waterfall* yang dimulai dari tahap *communication*, di mana kebutuhan fungsionalitas sistem didefinisikan. Selanjutnya, tahap *planning*, untuk menyusun jadwal pembuatan sistem. Tahap ketiga yaitu *modeling*, dilakukan perancangan dan pemodelan sistem melalui pembuatan *flowchart*, DFD, ERD, serta desain antarmuka. Tahap keempat adalah *construction* yang mencakup pengkodean program dan pengujian terhadap sistem yang telah dibangun. Tahap terakhir dari pembuatan sistem adalah *deployment* dengan melakukan konfigurasi *server* menggunakan metode *Continuous Integration/Continuous Deployment* (CI/CD) melalui *Github*.
- Hasil dan pembahasan, yaitu membahas hasil dari pembuatan sistem deteksi dini yang telah selesai, kemudian menuliskannya dalam sebuah laporan akhir.
- Kesimpulan dan saran, diambil dari hasil dan pembahasan penelitian yang telah dilakukan, maka akan ditarik kesimpulan, sedangkan saran didapatkan dari hasil pembuatan sistem yang telah dilakukan dan bersifat membangun untuk penelitian ke depannya.

3. HASIL DAN ANALISIS

3.1 Analisis Data

Pembuatan aturan klasifikasi mengikuti tahapan *Knowledge Discovery in Database* (KDD), mulai dari *data cleaning*, *data selection*, *data transformation*, *mining process*, dan terakhir *evaluation*. Proses *preprocessing* diawali dengan pengumpulan data rekam medis pasien menjadi satu *dataset*, kemudian dilakukan *data cleaning* dari *missing value*, duplikasi, dan inkonsistensi data. Selanjutnya, *data selection* dengan menghapus variabel yang tidak digunakan, ditransformasi ke bentuk numerik, serta dinormalisasi menggunakan *min-max normalization*. Tahap akhir adalah menentukan komposisi *dataset* dengan akurasi tertinggi.

3.1.1 Pengumpulan Data

Data penelitian dikumpulkan dari rekam medis pasien rawat inap di RSD dr. Soebandi Jember sebanyak 342 data, terdiri dari 246 diagnosis *Diabetes Mellitus* dan 96 diagnosis hipertensi. Pengumpulan data menggunakan lembar observasi berbentuk *checklist* dalam format *Excel* berdasarkan informasi pada rekam medis, yang selanjutnya digunakan pada tahap *preprocessing*.

Tabel 1. Hasil Pengumpulan Data

No.	Usia	JK	BB	TB	IMT	TD	GDA	RK	Diagnosa
1	63	P	63	156	26,3	140/80 mmHg	238 mg/dL	Ada	DM
2	43	L	90	180	27,7	128/88 mmHg	249 mg/dL		DM
3	64	P	50	152	21,6	118/86 mmHg	284 mg/dL		DM
4	74	L	60	147		143/93 mmHg	406 mg/dL	ada	DM
5	66	P	50	158		179/87 mmHg	397 mg/dL	Tidak ada	DM
...	...	..		...					
342	67	P	50	155	20	150/80 mmHg	254 mg/dL	Ada	DM

3.1.2 Data Cleaning

Tahap *preprocessing* selanjutnya adalah *data cleaning* untuk memastikan kualitas data. Proses ini meliputi penghapusan data dengan *missing value*, data duplikat, serta perbaikan data yang tidak konsisten. Tahapan ini menghasilkan 164 data bersih, terdiri dari 96 diagnosis *Diabetes Mellitus* dan 68 non *Diabetes Mellitus*.

Tabel 2. Hasil *Data Cleaning*

No.	Usia	JK	BB	TB	IMT	TD	GDA	RK	Diagnosa
1	≥ 40 tahun	P	63	156	Overweight	HT tingkat 1	Diabetes	Ada	DM
2	≥ 40 tahun	L	60	147	Overweight	HT tingkat 1	Diabetes	ada	DM
3	≥ 40 tahun	P	50	158	Normal	HT tingkat 2	Diabetes	Tidak ada	DM
4	≥ 40 tahun	L	68	165	Normal	HT tingkat 1	Diabetes	Tidak ada	DM
5	≥ 40 tahun	L	88	169	Obesitas	HT tingkat 2	Normal	Tidak ada	Non DM
...	...	.		...					
164	≥ 40 tahun	P	50	155	Normal	HT tingkat 1	Diabetes	Ada	DM

3.1.3 Data Selection

Tahap *data selection* dilakukan untuk memilih variabel yang digunakan dalam proses mining, dengan menghilangkan data yang tidak relevan [15]. Proses ini dilakukan menggunakan *Microsoft Excel* dengan meninjau dan menghapus variabel yang tidak digunakan, yaitu nomor rekam medis, berat badan, dan tinggi badan.

Tabel 3. Hasil *Data Selection*

No.	Usia	JK	IMT	TD	GDA	RK	Diagnosa
1	≥ 40 tahun	P	Overweight	HT tingkat 1	Diabetes	Ada	DM
2	≥ 40 tahun	L	Overweight	HT tingkat 1	Diabetes	ada	DM
3	≥ 40 tahun	P	Normal	HT tingkat 2	Diabetes	Tidak ada	DM
4	≥ 40 tahun	L	Normal	HT tingkat 1	Diabetes	Tidak ada	DM
5	≥ 40 tahun	L	Obesitas	HT tingkat 2	Normal	Tidak ada	Non DM
...	...	.					
164	≥ 40 tahun	P	Normal	HT tingkat 1	Diabetes	Ada	DM

3.1.4 Data Transformation

Tahap *preprocessing* terakhir adalah *data transformation*, yaitu mengubah variabel kategori menjadi bentuk numerik agar dapat diolah oleh algoritma melalui proses *encoding*. Metode yang digunakan adalah label *encoding*, yang mengonversi setiap nilai kategori menjadi numerik [16].

Tabel 4. Hasil *Data Transformation*

No.	Usia	JK	IMT	TD	GDA	RK	Diagnosa
1	1	1	2	2	1	1	DM
2	1	2	2	2	1	1	DM
3	1	1	3	1	1	2	DM
4	1	2	3	2	1	2	DM
5	1	2	1	1	3	2	Non DM
...	...	.					
164	1	1	3	2	1	1	DM

3.1.5 Normalisasi Data

Setelah *data transformation*, dilakukan normalisasi data menggunakan metode *min-max normalization* agar dataset siap digunakan pada algoritma *K-Nearest Neighbor* (K-NN). Normalisasi diperlukan karena K-NN menggunakan pengukuran jarak *Euclidean*, sehingga perbedaan skala antar fitur dapat memengaruhi hasil perhitungan jarak [17]. Metode *min-max normalization* digunakan untuk mengonversi data ke dalam rentang nilai baru guna menyeimbangkan skala antar fitur.

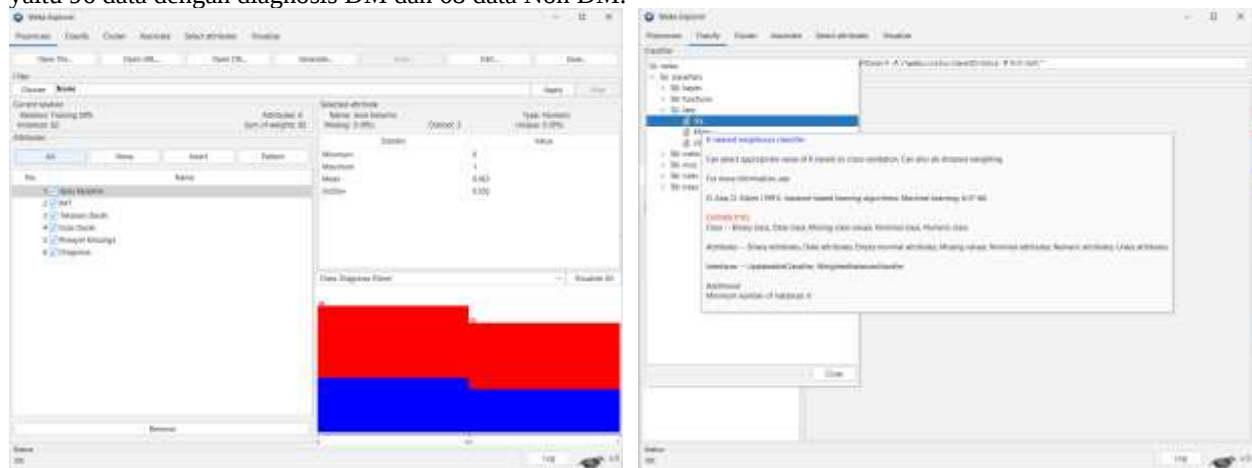
Tabel 5. Hasil Normalisasi Data Menggunakan *Min-Max Normalization*

No.	Usia	JK	IMT	TD	GDA	RK	Diagnosa
1	0.00	0.00	0.33	0.33	0.00	0.00	DM
2	0.00	1.00	0.33	0.33	0.00	0.00	DM

No.	Usia	JK	IMT	TD	GDA	RK	Diagnosa
3	0.00	0.00	0.67	0.00	0.00	1.00	DM
4	0.00	1.00	0.67	0.33	0.00	1.00	DM
5	0.00	1.00	0.00	0.00	1.00	1.00	Non DM
...	...	.					
164	0.00	0.00	0.67	0.33	0.00	0.00	DM

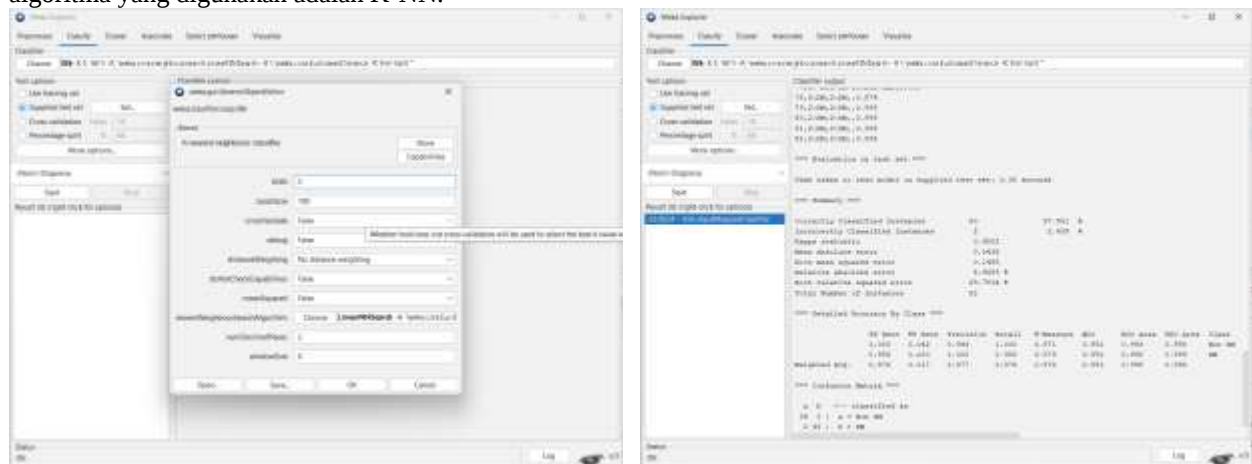
3.1.6 Mining Process

Proses *mining* dilakukan menggunakan WEKA dengan pembagian *data training* dan *data testing* menggunakan teknik *stratified sampling* pada rasio 90:10, 80:20, 70:30, 60:40, dan 50:50 untuk menjaga proporsi kelas dan mencegah bias dalam proses pelatihan serta pengujian model [16]. Dataset penelitian terdiri dari 164 data, yaitu 96 data dengan diagnosis DM dan 68 data Non DM.



Gambar 2. Tahapan Awal Mining Process

Gambar 2 merupakan tahapan awal *mining process*, dengan memilih file *data training* yang akan dilakukan perhitungan, lalu pilih atribut yang digunakan dan klik *classify*. Selanjutnya, pilih *classifiers* -> *lazy* -> *IBk*, karena algoritma yang digunakan adalah K-NN.



Gambar 3. Hasil Klasifikasi Algoritma K-NN

Gambar 3 menjelaskan langkah selanjutnya dengan pilih nilai K yang akan digunakan, kemudian akan muncul hasil klasifikasi yang dilakukan. Lakukan perhitungan berulang pada masing-masing nilai K untuk mendapatkan hasil nilai akurasi tertinggi.

Tabel 6. Perbandingan Hasil Perhitungan Akurasi Komposisi *Dataset*

Perbandingan Komposisi <i>Dataset</i>	Akurasi	Nilai K
90:10	94,12%	5, 7, 9, dan 11
80:20	93,94%	3, 5, 7, 9, dan 11
70:30	91,84%	3 dan 5
60:40	96,92%	3 dan 5
50:50	97,56%	3 dan 5

Hasil pengujian menunjukkan bahwa pembagian data training dan data testing 50:50 menghasilkan akurasi tertinggi sebesar 97,56%, sehingga digunakan dalam pembuatan aturan klasifikasi. Nilai K terbaik diperoleh pada K=3 dan K=5, sejalan dengan penelitian sebelumnya [18]. Dengan mempertimbangkan potensi *overfitting* pada nilai K yang terlalu kecil, nilai K yang digunakan dalam penelitian ini adalah K=5.

3.2 Pengujian Kinerja

Hasil penghitungan pada *mining process* menunjukkan bahwa akurasi tertinggi dalam pembentukan aturan klasifikasi diperoleh pada komposisi data training dan data testing 50:50. Perbedaan antara hasil klasifikasi dan diagnosis pada data rekam medis dievaluasi menggunakan *Confusion Matrix* untuk menilai kesesuaian hasil klasifikasi dengan data sebenarnya [19]. *Confusion Matrix* menampilkan jumlah klasifikasi benar dan salah serta menghasilkan nilai *accuracy*, *precision*, dan *recall*.

Perbandingan data aktual dan hasil klasifikasi menunjukkan bahwa terdapat 2 dari 82 data yang tidak sesuai dengan diagnosis sebenarnya, yaitu data dengan diagnosis DM yang diklasifikasikan sebagai Non DM. Kedua data tersebut dikategorikan sebagai *False Negative* (FN) pada *Confusion Matrix* yang disajikan pada Tabel 7.

Tabel 7. *Confusion Matrix* Hasil Klasifikasi WEKA

Aktual	Klasifikasi	
	DM	Non DM
DM	TP = 46	FN = 2
Non DM	FP = 0	TN = 34

Informasi yang diperoleh dari tabel *Confusion Matrix* juga dapat dimanfaatkan untuk menghitung nilai *precision* dan *recall*, sebagaimana ditunjukkan dalam penghitungan berikut.

$$Accuracy = \frac{TN+TP}{TP+FP+FN+TN} \times 100\% = \frac{34+46}{46+0+2+34} \times 100\% = \frac{80}{82} \times 100\% = 97,56\%$$

$$Precision = \frac{TP}{TP+FP} \times 100\% = \frac{46}{46+0} \times 100\% = \frac{46}{46} \times 100\% = 100\%$$

$$Recall = \frac{TP}{TP+FN} \times 100\% = \frac{46}{46+2} \times 100\% = \frac{46}{48} \times 100\% = 95,83\%$$

Berdasarkan *Confusion Matrix*, diperoleh nilai *accuracy* sebesar 97,56%, *precision* sebesar 100%, dan *recall* sebesar 95,83%. Nilai *accuracy* menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar, sementara *precision* sebesar 100% menandakan tidak terdapat kesalahan klasifikasi positif palsu. Nilai *recall* menunjukkan bahwa masih terdapat sebagian kecil kasus DM yang tidak terdeteksi. Secara keseluruhan, sistem memiliki performa yang sangat baik, namun tetap memerlukan penyempurnaan untuk meningkatkan sensitivitas dalam mendeteksi seluruh kasus DM.

3.3 Communication

Tahap *communication* dalam metode *Waterfall* diawali dengan pengumpulan kebutuhan fungsional dan non fungsional. Kebutuhan fungsional diperoleh melalui wawancara dengan dokter spesialis penyakit dalam, meliputi variabel faktor risiko DM (nama, usia, jenis kelamin, riwayat keluarga, berat badan, tinggi badan, IMT, tekanan darah, gula darah), serta kebutuhan pasien berupa informasi jenis DM, faktor risiko, gejala, pencegahan, dan rekomendasi penanganan. Sistem dirancang dengan fitur utama deteksi dini, riwayat deteksi, serta validasi dokter untuk meningkatkan akurasi data training. Kebutuhan non fungsional mencakup perangkat keras (laptop) dan perangkat lunak (*Windows 11*, *MS Office*, *Draw.io*, *Figma*, *Visual Studio Code*, *PHPmyAdmin*, *XAMPP*, *Github*). Dataset diambil dari rekam medis pasien rawat inap DM dan hipertensi di RS dr. Soebandi Jember, kemudian digunakan untuk membentuk aturan klasifikasi dengan algoritma K-NN.

3.4 Planning

Tahap berikutnya dalam proses perancangan sistem deteksi dini *Diabetes Mellitus* adalah *planning*. Pada tahap ini, dilakukan penyusunan jadwal kegiatan perancangan sistem yang divisualisasikan dalam bentuk *gant chart*, sebagaimana ditampilkan pada tabel 8.

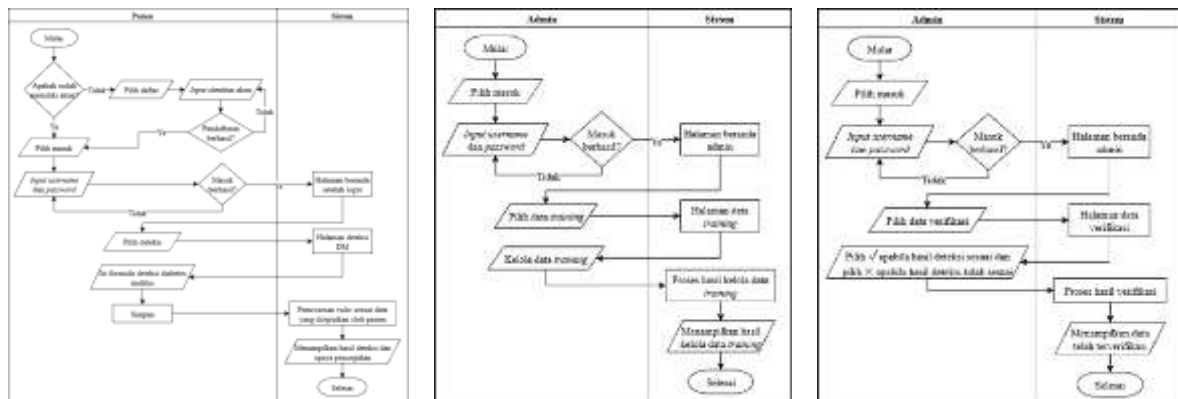
Tabel 8. *Gantt Chart* Perancangan Sistem Deteksi Dini Diabetes Melitus

Kegiatan	Bulan ke-1				Bulan ke-2					Bulan ke-3				Bulan ke-4		
	1	2	3	4	1	2	3	4	5	1	2	3	4	1	2	3
Pembuatan desain antarmuka, <i>Flowchart</i> , <i>Data Flow Diagram</i> (DFD), dan <i>Entity Relationship Diagram</i> (ERD).																
Pembuatan desain <i>database</i>																
Implementasi antarmuka ke dalam bahasa pemrograman																
Pengujian fungsionalitas dan kinerja sistem deteksi dini penyakit <i>Diabetes Mellitus</i> dengan <i>black box testing</i>																

3.5 Modeling

3.5.1 Flowchart

a. Flowchart pasien

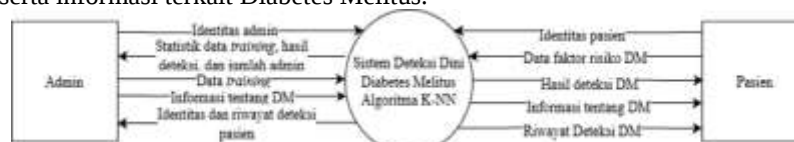


Gambar 4. Flowchart Pasien dan Admin

Gambar *flowchart* pasien menggambarkan alur kerja sistem deteksi dini *Diabetes Mellitus* dari sisi pasien. Proses dimulai dengan login atau pendaftaran akun, kemudian pasien mengakses menu deteksi dan mengisi formulir faktor risiko. Data yang dikirimkan diproses oleh sistem berdasarkan aturan yang telah ditentukan, selanjutnya sistem menampilkan hasil deteksi serta upaya pencegahan sesuai dengan hasil klasifikasi. Gambar *flowchart* admin pertama menunjukkan alur kerja sistem deteksi dini *Diabetes Mellitus* pada pengelolaan *data training* oleh admin. Proses dimulai dengan login admin, dilanjutkan pengelolaan *data training* melalui menu yang tersedia, meliputi melihat, menambah, mengedit, mengimpor, dan menghapus data. Selanjutnya, sistem memproses dan menampilkan hasil pengelolaan *data training*. Gambar *flowchart* admin kedua menunjukkan alur kerja sistem deteksi dini *Diabetes Mellitus* pada proses verifikasi hasil deteksi pasien oleh admin. Proses dimulai dengan login admin, dilanjutkan pemeriksaan dan verifikasi hasil deteksi melalui menu data verifikasi. Admin memberikan tanda ✓ atau ✗ sesuai hasil pemeriksaan, kemudian sistem memproses dan menampilkan data yang telah terverifikasi.

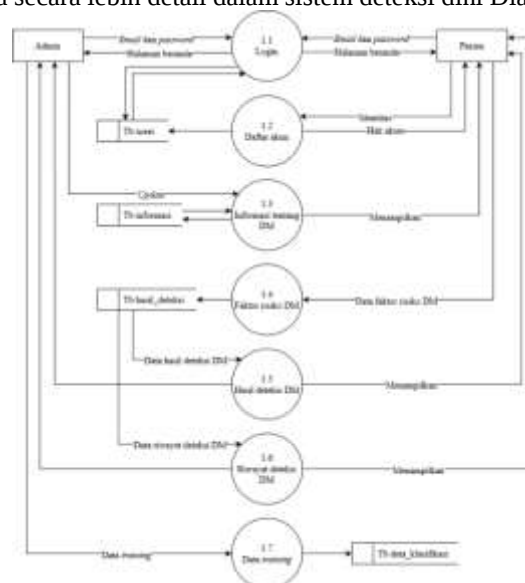
3.5.2 Data Flow Diagram (DFD)

Perancangan sistem deteksi dini *Diabetes Mellitus* menggunakan *Data Flow Diagram* (DFD) level 0 dan level 1. DFD level 0 menggambarkan aliran data secara umum dalam sistem yang melibatkan dua entitas, yaitu pasien dan admin. Pasien menginput identitas dan data faktor risiko untuk diproses oleh sistem sehingga menghasilkan prediksi, serta dapat melihat informasi dan riwayat deteksi. Admin berperan dalam mengelola *data training*, identitas dan riwayat pasien, serta informasi terkait *Diabetes Mellitus*.



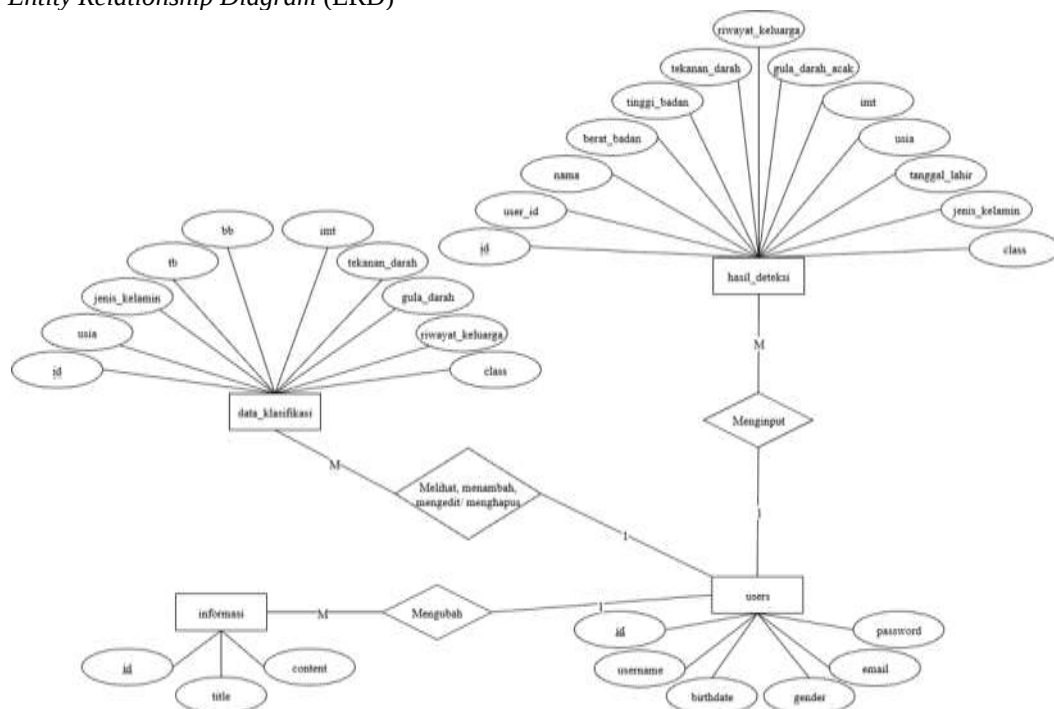
Gambar 5. Data Flow Diagram Level 0

Data Flow Diagram (DFD) level 1 merupakan penjabaran lebih rinci dari proses utama pada DFD level 0 yang menggambarkan aliran data secara lebih detail dalam sistem deteksi dini *Diabetes Mellitus*.



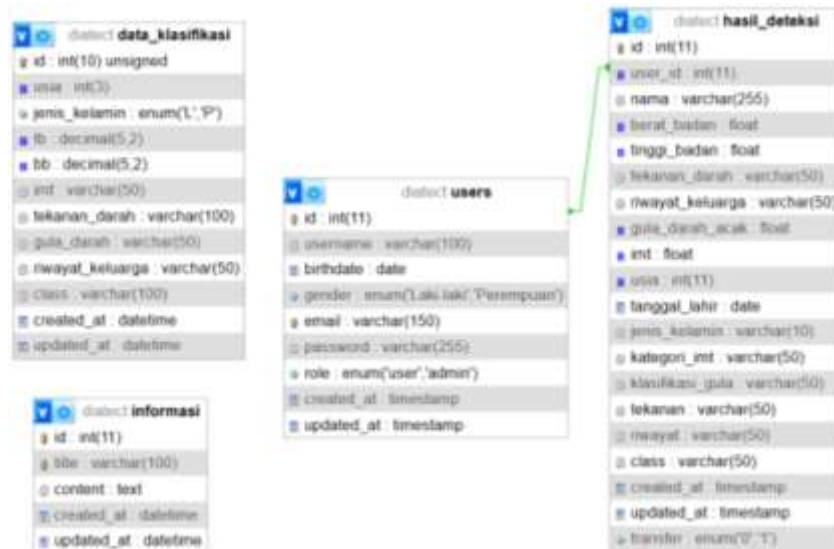
Gambar 6. Data Flow Diagram Level 1

3.5.3 Entity Relationship Diagram (ERD)



Gambar 7. Entity Relationship Diagram

Entity Relationship Diagram (ERD) sistem deteksi dini *Diabetes Mellitus* ditunjukkan pada Gambar 7 yang terdiri dari empat entitas, yaitu *users*, *informasi*, *data_klasifikasi*, dan *hasil_deteksi*. Pasien menginput data identitas dan faktor risiko yang disimpan pada tabel *hasil_deteksi*, sedangkan admin mengelola informasi DM dan data training yang disimpan pada tabel *informasi* dan *data_klasifikasi*. Detail struktur basis data ditampilkan pada *Physical Data Model* (PDM) pada Gambar 8.



Gambar 8. Physical Data Model

3.5.4 Design Interface

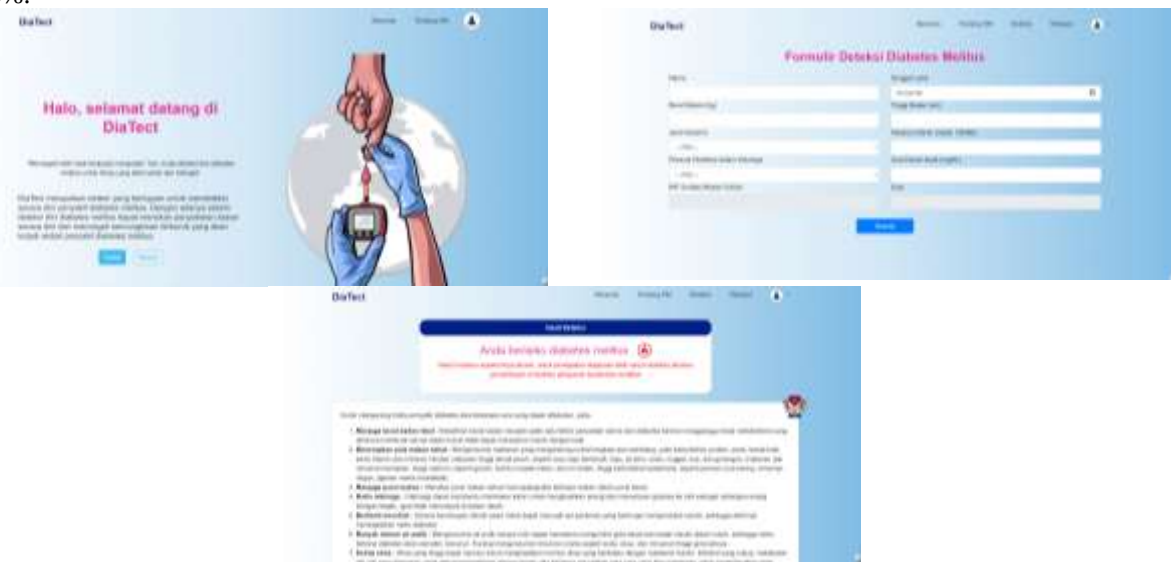
Design Interface sistem deteksi dini *Diabetes Mellitus* pasien mencakup halaman beranda sebelum dan sesudah *login*, halaman registrasi dan *login*, profil pengguna, informasi *Diabetes Mellitus*, deteksi DM, hasil deteksi, riwayat deteksi, serta *logout*. Halaman beranda memberikan informasi umum sistem, sementara setelah *login* pengguna dapat mengakses menu deteksi dan riwayat. Proses registrasi dan *login* digunakan untuk pengelolaan akun pengguna, sedangkan halaman profil memungkinkan pembaruan data. Halaman informasi DM menyajikan edukasi terkait *Diabetes Mellitus*. Halaman deteksi berisi formulir pengisian faktor risiko, yang selanjutnya menghasilkan tampilan hasil deteksi berisiko atau tidak berisiko DM beserta rekomendasi. Riwayat deteksi menampilkan hasil deteksi yang telah dilakukan, dan halaman *logout* digunakan untuk keluar dari sistem.

Design Interface admin pada sistem deteksi dini *Diabetes Mellitus* mencakup halaman beranda admin setelah *login* yang menampilkan statistik *data training*, hasil deteksi, dan jumlah admin. Admin dapat mengakses halaman hasil deteksi untuk melihat riwayat deteksi pengguna, mengelola data training secara manual maupun melalui unggah

file Excel, serta mengedit informasi mengenai Diabetes Melitus. Selain itu, admin dapat mengelola data admin dan pasien, termasuk melihat, menambah, mengubah, dan menghapus data serta riwayat deteksi. Sistem juga menyediakan halaman verifikasi hasil deteksi pasien, di mana data yang telah diverifikasi akan otomatis masuk ke dalam *data training*.

3.6 Construction

Tahap construction dalam pengembangan sistem deteksi dini *Diabetes Mellitus* meliputi proses pengkodean dan pengujian sistem. Pengkodean dilakukan menggunakan *Visual Studio Code* dengan bahasa pemrograman PHP, database *MySQL*, dan *framework CodeIgniter 3*. Pengujian sistem dilakukan melalui pengujian fungsionalitas menggunakan metode *blackbox testing* serta pengujian kinerja dengan *82 data testing*. Hasil pengujian menunjukkan bahwa kinerja sistem sesuai dengan aplikasi WEKA, dengan dua data tidak sesuai diagnosis dan nilai *accuracy* sebesar 97,56%.



Gambar 9. Sistem Deteksi Dini *Diabetes Mellitus* (DiaTect)

Pada gambar di atas merupakan tampilan sistem deteksi dini *Diabetes Mellitus*, gambar pertama merupakan halaman awal sistem, selanjutnya tampilan menu utama sistem deteksi dini *Diabetes Mellitus* yang berisi formulir deteksi yang digunakan untuk melakukan deteksi dini. Setelah mengisi formulir deteksi, maka hasil deteksi akan di tampilkan langsung setelah pengguna menekan submit.

3.7 Deployment

Tahap *deployment* adalah implementasi sistem ke pengguna melalui konfigurasi, pemeliharaan, evaluasi, dan pengembangan lebih lanjut berdasarkan umpan balik. Namun, penelitian ini hanya melakukan konfigurasi *server* dengan metode CI/CD melalui *Github* agar *website* dapat dipublikasikan dan diakses secara daring oleh pengguna. Sistem deteksi dini DM dapat diakses melalui link <https://diatetect.mikpolije.com/>.

4. KESIMPULAN

Penelitian ini menggunakan algoritma *K-Nearest Neighbor* (K-NN) untuk deteksi dini *Diabetes Mellitus* berdasarkan rekam medis pasien rawat inap di RSD dr. Soebandi Jember. Dari 342 data awal diperoleh 164 data bersih melalui *preprocessing*. Data kemudian dibagi dengan rasio 50:50, yaitu 50% sebagai *data training* dan 50% sebagai *data testing*. Nilai K=5 dipilih karena menghasilkan performa terbaik dibandingkan nilai K lainnya, seperti K=7 dan K=9, yang menunjukkan penurunan nilai akurasi dan kestabilan klasifikasi. *Dataset* mencakup 6 variabel faktor risiko DM yaitu riwayat penyakit *Diabetes Mellitus* dalam keluarga, usia, jenis kelamin, kadar gula darah acak, tekanan darah, serta indeks massa tubuh (IMT).

Pengujian kinerja dilakukan menggunakan *Confusion Matrix* untuk menghitung *accuracy*, *precision*, dan *recall*. Hasil menunjukkan *accuracy* sebesar 97,56%, yang berarti model mampu mengklasifikasikan data dengan tingkat ketepatan tinggi. Nilai *precision* 100% menunjukkan seluruh data yang diprediksi positif benar-benar positif, sedangkan *recall* 95,83% menunjukkan sebagian besar kasus positif berhasil terdeteksi oleh sistem. Sistem dikembangkan dengan metode *Waterfall* dan diimplementasikan berbasis *website* menggunakan *PHP-MySQL*, dengan *framework CodeIgniter 3*, diuji melalui *blackbox testing*, divalidasi dengan WEKA, serta dipublikasikan secara daring melalui konfigurasi *server* CI/CD berbasis *GitHub*.

REFERENSI

- [1] Depkes RI, "Pedoman Pengendalian Diabetes Melitus dan Penyakit Metabolik." p. 82, 2008. [Online]. Available: file:///D:/ebook/dsa664.pdf
- [2] World Health Organization, "10 Penyebab Kematian Teratas," *World Health Organization*, 2020.

- https://www-who-int.translate.goog/news-room/fact-sheets/detail/the-top-10-causes-of-death?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc (accessed Sep. 22, 2023).
- [3] Ditpui, "Diabetes Penyebab Kematian Tertinggi di Indonesia: Batasi dengan Snack Sehat Rendah Gula," *Direktorat Pengembangan Usaha Universitas Gajah Mada*, 2023. <https://ditpui.ugm.ac.id/diabetes-penyebab-kematian-tertinggi-di-indonesia-batasi-dengan-snack-sehat-rendah-gula/> (accessed Sep. 22, 2023).
- [4] Kemenkes RI, "Hasil Riset Kesehatan Dasar Tahun 2018," *Kementrian Kesehat. RI*, vol. 53, no. 9, pp. 1689–1699, 2018.
- [5] Dinkes Jatim, *Profil Kesehatan 2021*. 2021. doi: 10.21831/dinamika.v3i1.19144.
- [6] R. Pahlevi, "Kasus Kematian Akibat Diabetes di Indonesia Terbesar Keenam di Dunia," *databoks*, 2021. <https://databoks.katadata.co.id/datapublish/2021/11/26/kasus-kematian-akibat-diabetes-di-indonesia-terbesar-keenam-di-dunia> (accessed Sep. 22, 2023).
- [7] R. P. Kurniadi, R. R. Saedudin, and V. P. Widartha, "Perbandingan Akurasi Algoritma K-Nearest Neighbor Dan Logistic Regression Untuk Klasifikasi Penyakit Diabetes," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9757–9764, 2021.
- [8] H. A. Dwi Fasnuari, H. Yuana, and M. T. Chulkamdi, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus," *J. Ilm. Tek. Inform.*, vol. 16, no. 2, pp. 133–142, 2022, doi: 10.35457/antivirus.v16i2.2445.
- [9] I. L. F. Amien, W. Astuti, and K. M. Lhaksamana, "Perbandingan Metode Naïve Bayes dan KNN (K-Nearest Neighbor) dalam Klasifikasi Penyakit Diabetes," *e-Proceeding Eng.*, vol. 10, no. 2, pp. 1911–1920, 2023.
- [10] F. Tolana, I. Muzakkir, and A. Riadi, "Metode K-Nearest Neighbor (KNN) Untuk Sistem Pakar Diagnosa Penyakit Pneumonia Pada Balita," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 6, no. 3, pp. 363–373, 2023.
- [11] S. K. P. Loka and A. Marsal, "Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier untuk Klasifikasi Status Gizi Pada Balita," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. April, pp. 8–14, 2023, [Online]. Available: <https://journal.irpi.or.id/index.php/malcom/article/view/474>
- [12] A. H. Blasi, "What is The Mininum Amount of Data Needed When Carrying Out Analysis Via Datamining Techniques Like Decision Tree?," *ResearchGate*, 2020. <https://www.researchgate.net/post/What-is-the-mininum-amount-of-data-needed-when-carrying-out-analysis-via-datamining-techniques-like-decision-tree>
- [13] P. N. Lubis, N. A. Pasaribu, N. Meilani, and M. S. Hasibuan, "Implementasi Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Memonitoring Data Pelanggan Oleh PT. Telkom Akses Pematangsiantar," *J. Ilmu Komput. dan Bisnis*, vol. 14, no. 1, pp. 220–227, 2023, doi: 10.47927/jikb.v14i1.473.
- [14] S. Romlah and Z. Fatah, "Penerapan Data Mining Untuk Memprediksi Hafalan Santri Tahfidz Menggunakan Metode Decision Tree (Studi Kasus MIK Sarina)," *Gudang J. Multidisiplin Ilmu*, vol. 2, no. 10, pp. 160–166, 2024.
- [15] D. Retnosari, "Sistem Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa," *J. Integr. Sist. Ind. UMJ*, vol. 1, no. 2, pp. 13–20, 2014.
- [16] J. K. Wororomi et al., *Data Mining (Memahami Pola di Balik Angka)*. CV. Eureka Media Aksara, 2024.
- [17] P. P. Allorerung, A. Erna, and M. Bagussahrir, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," *J. Inform. Sunan Kalijaga*, vol. 9, no. 3, pp. 178–191, 2024.
- [18] A. Ananda, E. Sulistianingsih, and Yundari, "Metode Ensemble K-Nearest Neighbor Untuk Peningkatan Akurasi Prediksi Indeks Harga Saham Gabungan Di Indonesia," *Bul. Ilm. Math. Stat. dan Ter.*, vol. 13, no. 3, pp. 339–348, 2024.
- [19] R. Nurhidayat and K. E. Dewi, "Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek," *Komputa J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, pp. 91–100, 2023, doi: 10.34010/komputa.v12i1.9458.