

CEK TURNITIN

SKRIPSI KHAIRUNNISA SI

 SATSET DIGITAL 1

Document Details

Submission ID

trn:oid:::3618:150982425

Submission Date

Sep 8, 2026, 3:24 PM GMT+7

Download Date

Sep 8, 2026, 3:32 PM GMT+7

File Name

SKRIPSI KHAIRUNNISA SI.pdf

File Size

1.6 MB

63 Pages

8,364 Words

60,357 Characters

36% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Top Sources

- 26%  Internet sources
- 15%  Publications
- 30%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

- 26% Internet sources
- 15% Publications
- 30% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	repository.unisi.ac.id	6%
2	Student papers	UPN Veteran Yogyakarta on 2026-07-06	<1%
3	Internet	teknosi.fti.unand.ac.id	<1%
4	Student papers	Universitas Khairun on 2026-05-25	<1%
5	Student papers	Universitas Sangga Buana YPKP on 2026-08-03	<1%
6	Student papers	Universitas Muhammadiyah Makassar on 2026-06-25	<1%
7	Student papers	Universitas Sangga Buana YPKP on 2026-07-23	<1%
8	Internet	repository.uin-suska.ac.id	<1%
9	Internet	eprints.amikom.ac.id	<1%
10	Internet	inacl.id	<1%
11	Internet	repository.ub.ac.id	<1%

12	Internet	ejournal.unesa.ac.id	<1%
13	Internet	repository.ittelkom-pwt.ac.id	<1%
14	Student papers	Universitas Sangga Buana YPKP on 2026-07-29	<1%
15	Student papers	Universitas Putera Batam on 2026-07-24	<1%
16	Internet	jurnal.ulb.ac.id	<1%
17	Internet	ejournal.itn.ac.id	<1%
18	Internet	positori.unwira.ac.id	<1%
19	Student papers	Konsorsium PTS Batch 5 on 2026-06-26	<1%
20	Internet	core.ac.uk	<1%
21	Publication	Panji Pangestu Saputra, Hasbi Firmansyah, Rizki Prasetyo Tulodo, Priyo Haryok...	<1%
22	Internet	ejournal.uhb.ac.id	<1%
23	Internet	repository.umsu.ac.id	<1%
24	Student papers	Universitas Pendidikan Ganesha on 2026-03-03	<1%
25	Internet	docplayer.info	<1%

26	Internet	journal.stiestekom.ac.id	<1%
27	Internet	Widiastiwi, Yuni, Ernawati, Iin. "Klasifikasi Penyakit Batu Ginjal Menggunakan Alg...	<1%
28	Publication	Fidalina Nirigi, Mochammad Triyanto, Mohammad Rezza Pahlevi, Athia Saelan, Fa...	<1%
29	Student papers	Universitas Sanata Dharma on 2026-06-10	<1%
30	Student papers	Universitas Sebelas Maret on 2025-12-13	<1%
31	Internet	eprints.stikesmhk.ac.id	<1%
32	Internet	repositori.pustaka.serambimekkah.ac.id	<1%
33	Student papers	Universitas Sangga Buana YPKP on 2026-07-28	<1%
34	Internet	jurnal.stikeskesosi.ac.id	<1%
35	Internet	ojs.uma.ac.id	<1%
36	Student papers	UIN Sjech M.Djamil Djambek Bukittinggi on 2026-07-08	<1%
37	Student papers	Universitas Brawijaya on 2017-06-16	<1%
38	Student papers	Universitas Islam Sumatera Utara on 2026-04-20	<1%
39	Internet	repository.maranatha.edu	<1%

40	Publication	Halimukid Halimukid. "ANALISIS KELULUSAN SISWA MENGGUNAKAN METODE KN...	<1%
41	Student papers	Universitas Islam Indonesia on 2018-07-21	<1%
42	Student papers	Universitas Merdeka Malang on 2025-11-24	<1%
43	Publication	M Reza Kausar, Wahyu Fuadi, Zahratul Fitri. "Classification of Coronary Heart Dise...	<1%
44	Student papers	UPN Veteran Yogyakarta on 2026-06-26	<1%
45	Student papers	Universitas Islam Riau on 2025-12-10	<1%
46	Internet	ejournal.pnc.ac.id	<1%
47	Internet	jurnal.alimspublishing.co.id	<1%
48	Publication	Dwito Julian Tambunan, Agustina Simbolon, Sri Ulina Br Marmata, Sardo Parding...	<1%
49	Student papers	UIN Raden Intan Lampung on 2026-01-22	<1%
50	Student papers	Universitas Airlangga-1 on 2025-12-15	<1%
51	Student papers	Universitas Pancasila on 2026-09-02	<1%
52	Student papers	Universitas Sangga Buana YPKP on 2026-07-22	<1%
53	Publication	Yeheskiel Ravena Damanik, Albert Adolf Angie MS, Esther Laura Rumahorbo, Elis...	<1%

54	Internet	digilib.unila.ac.id	<1%
55	Student papers	Universitas Brawijaya on 2019-01-13	<1%
56	Student papers	Universitas Islam Riau on 2026-08-18	<1%
57	Internet	bajangjournal.com	<1%
58	Internet	repository.teknokrat.ac.id	<1%
59	Internet	repository.unissula.ac.id	<1%
60	Internet	repository.upnjatim.ac.id	<1%
61	Internet	repository.upstegal.ac.id	<1%
62	Student papers	Universitas Brawijaya on 2017-07-31	<1%
63	Student papers	Universitas Islam Riau on 2026-06-15	<1%
64	Student papers	Universitas Muhammadiyah Jember on 2026-03-01	<1%
65	Student papers	Universitas Tarumanagara on 2025-12-14	<1%
66	Student papers	President University on 2015-04-13	<1%
67	Internet	jurnal.lenteranusa.id	<1%

68	Internet	munakab.go.id	<1%
69	Internet	rama.uniku.ac.id	<1%
70	Publication	Larisa Oriana, Anisa Dwi Sanggar Wati, Anggi Putri Ramahdani, Natasya Nurul Sa...	<1%
71	Student papers	Universitas Negeri Padang on 2025-09-12	<1%
72	Student papers	Universitas Pamulang on 2026-07-28	<1%
73	Student papers	Universitas Sanata Dharma on 2026-06-10	<1%
74	Student papers	Universitas Sangga Buana YPKP on 2026-08-31	<1%
75	Internet	digilib.uinsa.ac.id	<1%
76	Internet	dspace.uui.ac.id	<1%
77	Internet	jurnal.stmikmethodistbinjai.ac.id	<1%
78	Internet	jurnal.umk.ac.id	<1%
79	Internet	repository.upi.edu	<1%
80	Internet	123dok.com	<1%
81	Publication	Dian Megah Sari, Chairi Nur Insani, Adi Heri, Nurhikma Arifin. "Penentuan Takara...	<1%

82	Student papers	UNIVERSITAS BUDI LUHUR on 2025-07-24	<1%
83	Student papers	Universitas Brawijaya on 2018-01-08	<1%
84	Student papers	Universitas Muria Kudus on 2016-04-06	<1%
85	Student papers	Universitas Pamulang on 2026-05-04	<1%
86	Internet	id.123dok.com	<1%
87	Internet	journal.unpas.ac.id	<1%
88	Internet	opac.uad.ac.id	<1%
89	Student papers	Sriwijaya University on 2023-06-20	<1%
90	Student papers	Universitas Brawijaya on 2017-07-08	<1%
91	Student papers	Universitas Sanata Dharma on 2026-06-10	<1%
92	Internet	fisip.uajy.ac.id	<1%
93	Internet	kc.umh.ac.id	<1%
94	Internet	www.jurnal.stie-aas.ac.id	<1%
95	Student papers	Charles University on 2024-08-01	<1%

96	Student papers	Universitas Bengkulu on 2025-10-17	<1%
97	Student papers	Universitas Brawijaya on 2018-07-05	<1%
98	Student papers	Universitas Pendidikan Indonesia on 2014-05-30	<1%
99	Student papers	Universitas Sangga Buana YPKP on 2026-07-29	<1%
100	Internet	adoc.pub	<1%
101	Internet	ejurnal.provisi.ac.id	<1%
102	Internet	garuda.ristekbrin.go.id	<1%
103	Internet	github.com	<1%
104	Internet	informatika.universitasdumai.ac.id	<1%
105	Internet	journal.ubpkarawang.ac.id	<1%
106	Internet	jurnalfkip.samawa-university.ac.id	<1%
107	Internet	jutif.if.unsoed.ac.id	<1%
108	Publication	Bhabani Prasanna Pattanaik, Dharendra Kumar Shukla, Sambit Satpathy, Kamala...	<1%
109	Publication	Citha Nastitie, Irma Handayani2. "PERBANDINGAN METODE MOBILENETV2 DAN E...	<1%

110	Student papers	UNIVERSITAS BUDI LUHUR on 2026-07-13	<1%
111	Student papers	Universitas Brawijaya on 2019-08-28	<1%
112	Student papers	Universitas Dian Nuswantoro on 2020-10-29	<1%
113	Student papers	Universitas Muhammadiyah Palembang on 2026-06-10	<1%
114	Student papers	Universitas Pamulang on 2026-07-03	<1%
115	Student papers	Universitas Pamulang on 2026-08-05	<1%
116	Student papers	Universitas Sangga Buana YPKP on 2026-07-31	<1%
117	Student papers	Universitas Sangga Buana YPKP on 2026-07-31	<1%
118	Internet	blog.ub.ac.id	<1%
119	Internet	e-journal.hamzanwadi.ac.id	<1%
120	Internet	ejurnal.stmik-budidarma.ac.id	<1%
121	Internet	jurnal.bsi.ac.id	<1%
122	Internet	jurnal.polbeng.ac.id	<1%
123	Internet	jurnal.uimedan.ac.id	<1%

124	Internet	publikasi.dinus.ac.id	<1%
125	Internet	rama.unimal.ac.id	<1%
126	Internet	repository.ar-raniry.ac.id	<1%
127	Internet	repository.uncp.ac.id	<1%
128	Internet	repository.unilak.ac.id	<1%
129	Student papers	Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur II on 2025-08-20	<1%
130	Student papers	Sriwijaya University on 2024-05-02	<1%
131	Student papers	Sriwijaya University on 2025-11-24	<1%
132	Student papers	UPN Veteran Yogyakarta on 2026-05-04	<1%
133	Student papers	Universitas Amikom on 2023-06-06	<1%
134	Student papers	Universitas Brawijaya on 2018-12-18	<1%
135	Student papers	Universitas Negeri Semarang on 2025-11-17	<1%
136	Student papers	Universitas Pembangunan Panca Budi on 2026-08-02	<1%
137	Student papers	Ajou University Graduate School on 2024-07-01	<1%

138	Publication	Arimbi Puspitasari, Diana Sava Salsabila, Dwi Roliawati. "Penerapan ResNet-50 C...	<1%
139	Publication	Fastabiqul Khusna, Khothibul Umam, Siti Nur'aini, Maya Rini Handayani. "EVAL...	<1%
140	Publication	Ikka Ayu Elfutrianna, Rastri Prathivi. "Komparasi Metode SVM dan Adaboost untu...	<1%
141	Student papers	UPN Veteran Yogyakarta on 2026-07-21	<1%
142	Student papers	Universitas Brawijaya on 2018-12-12	<1%
143	Student papers	Universitas Merdeka Malang on 2018-04-11	<1%
144	Student papers	iGroup on 2011-04-20	<1%

23

**KLASIFIKASI PENYAKIT PASIEN RAWAT JALAN
MENGUNAKAN METODE *K-NEAREST NEIGHBORS*
(KNN)**

SKRIPSI



Disusun Oleh

KHAIRUNNISA

403221010019

1

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS ISLAM INDRAGIRI
TEMBILAHAN**

2025

PERSETUJUAN

PENGESAHAN

SURAT PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama Mahasiswa : **KHAIRUNNISA**
NIM : 403221010019
Program Studi : Sistem Informasi

Menyatakan bahwa skripsi sbb:

Judul bahasa Indonesia :

KLASIFIKASI PENYAKIT PASIEN RAWAT JALAN MENGGUNAKAN
METODE K-NEAREST NEIGHBORS (KNN)

Judul bahasa Inggris :

*CLASSIFICATION OF OUTPATIENT DISEASES USING THE K-NEAREST
NEIGHBORS (KNN) METHOD*

Pembimbing Utama : Dr. Bayu Rianto, S.SI., M.Kom

Pembimbing Pendamping : Muh. Rasyid Ridha, S.SI., M.Kom

1. Laporan Tugas Akhir ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas Islam Indragiri maupun di perguruan tinggi lainnya;
2. Laporan Tugas Akhir ini merupakan gagasan, rumusan dan pemikiran SAYA sendiri, tanpa bantuan pihak lain kecuali arahan Tim Dosen Pembimbing;
3. Dalam laporan Tugas Akhir ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada laporan ini;
4. Perangkat lunak yang digunakan dalam Tugas Akhir ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas Islam Indragiri;
5. Pernyataan ini SAYA buat dengan sesungguhnya apabila di kemudian hari terdapat penyimpangan dan ketidak benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku diperguruan tinggi.

Tembilahan, 8 Juni 2026

Yang Menyatakan,

Materai

KHAIRUNNISA

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala limpahan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyusun proposal penelitian yang berjudul “Klasifikasi Penyakit Pasien Rawat Jalan di Puskesmas Menggunakan *K-Nearest Neighbors* (KNN)”. Proposal ini disusun sebagai langkah awal dalam pelaksanaan penelitian yang bertujuan untuk mendukung peningkatan kualitas pengelolaan data kesehatan di Puskesmas melalui penerapan teknologi informasi dan metode data mining.

Penulis sudah berupaya semaksimal mungkin dalam menyelesaikan proposal ini. Namun, penulis menyadari bahwa didalam proposal ini tentunya masih terdapat kekurangan-kekurangan, baik dalam penyajian materi ataupun dalam materi itu sendiri. Dalam pembuatan proposal ini tidak lepas dari berbagai pihak yang telah membantu dari segi material atau spiritual. Atas segala bimbingan, dorongan dan bantuan secara langsung maupun tidak langsung yang telah diberikan

Pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada yang terhormat :

1. Kedua orang tua saya
2. Bapak Dr. Najamuddin Lc., MA, Sebagai Rektor Universitas Islam Indragiri
3. Bapak Dr. Bayu Rianto, S.SI., M.Kom Sebagai Dekan Fakultas Teknik dan Ilmu Komputer Universitas Islam Indragiri sekaligus Dosen Pembimbing yang telah memberi bimbingan selama penulisan proposal ini

4. Ibu Fitri Yunita S.SI.,M.Kom Sebagai Ketua Prodi Program Studi Sistem Informasi Fakultas Teknik dan Ilmu Komputer
5. Bapak Muh.Rasyid Ridha, S.SI, M.Kom Sebagai Dosen Pembimbing pendamping yang telah memberi bimbingan selama penulisan proposal ini
6. Seluruh keluarga besar dan seluruh teman teman saya yang telah memberikan dukungan selama ini.

Semoga proposal penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang kesehatan dan teknologi informasi, serta menjadi langkah awal yang baik dalam pelaksanaan penelitian selanjutnya.

Tembilahan, November 2025

khairunnisa

DAFTAR ISI

PERSETUJUAN	i
PENGESAHAN	ii
DAFTAR ISI.....	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
INTISARI	xi
ABSTRACT	xii
BAB I.....	1
PENDAHULUAN	1
1.1Latar Belakang	1
1.2Rumusan Masalah	3
1.3Batasan Masalah.....	4
1.4Tujuan Penelitian	4
1.5Manfaat Penelitian	4
1.6Sistematika Penulisan.....	5
BAB II	6
TINJAUAN LITERATUR.....	6
2.1 Pendahuluan	6
2.2 tinjauan literature	6
2.3 Rangkuman	14
BAB III	16
METODE PENELITIAN	16
3.1Pendahuluan	16
3.2kerangka penelitian	16
3.3Data Pasien Rawat Jalan.....	17
3.4Preprocessing Data.....	17
3.5Normalisasi Data	19

3.6Pembagian Data	20
3.7Proses KNN	20
3.8Penentuan K Tetangga Terdekat.....	22
3.9Evaluasi Model	23
Hasil & kesimpulan.....	24
BAB IV	26
HASIL DAN PEMBAHASAN	26
4.1 Perbandingan Akurasi Antar Subset.....	26
4.2 Hasil Evaluasi lengkap pada Data Test	27
4.3 Performa per Penyakit	27
4.4 Fitur dan Target Klasifikasi.....	29
4.5 Kelas Penyakit.....	30
4.6 Tahapan Prapocessing	31
4.7 Metode Klasifikasi.....	32
4.8 Penanganan Ketidakseimbangan Data.....	33
4.9 Evaluasi Model.....	33
4.9 Implementasi Sistem	38
4.10. Tampilan Aplikasi	39
a.Tampilan awal	39
b.Login admin	41
c.Dashboard admin	42
BAB V	45
Kesimpulan dan Saran.....	45
DAFTAR PUSTAKA	47

DAFTAR TABEL

Tabel 2.2 Penelitian terdahulu.....	6
Tabel 3.1Pembagian data dan fungsinya.....	21
Tabel 3.2 Akurasi train dan validation untuk tiap nilai k.....	23
Tabel 4.1 Perbandingan akurasi model final antar subset.....	27
Tabel 4.2 Ringkasan file dan struktur dataset	28
Tabel 4.3 Ringkasan dataset penelitian	28
Tabel 4.4 Kelas penyakit.....	29
Tabel 4.5 Hasil model terbaik	30
Tabel 4.6 Fitur dan target yang digunakan dalam penelitian	31
Tabel 4.7 Distribusi kelas penyakit pada dataset final	32
Tabel 4.8 Contoh normalisasi istilah pada teks gejala	34
Tabel 4.9 Konfigurasi model klasifikasi	34
Tabel 4.10 Strategi balancing data	35
Tabel 4.11 Hasil evaluasi pada beberapa nilai K	36

DAFTAR GAMBAR

Gambar 1. Kerangka penelitian.....	17
Gambar 3. Karva akurasi train vs validation terhadap nilai K.....	24
Gambar 4. Counfusion matrix.....	37
Gambar 5. Grafik Performa model per nilai K	39
Gambar 6. Tampilan website	43
Gambar 7. Login admin	44
Gambar 8. Dashboard admin.....	46

INTISARI

Penelitian ini mengembangkan sistem klasifikasi diagnosis penyakit pasien rawat jalan menggunakan metode K-Nearest Neighbors (KNN) berdasarkan gejala pasien. Data diproses melalui tahap Preprocessing, ekstraksi fitur menggunakan TF-IDF, dan klasifikasi menggunakan Cosine Distance. Sistem diimplementasikan dalam aplikasi berbasis web dan menghasilkan performa terbaik K=7 dengan accuracy 94.67%, Macro F1 94.54%, dan Top-5 Accuracy 98%. Hasil penelitian menunjukkan bahwa metode KNN mampu membantu memberikan rekomendasi diagnosis awal secara efektif.

ABSTRACT

Penelitian ini bertujuan membangun sistem klasifikasi diagnosis penyakit pasien rawat jalan menggunakan metode K-Nearest Neighbors (KNN) berdasarkan gejala pasien. Dataset yang digunakan berasal dari data rekam medis pasien yang diproses melalui tahap *preprocessing*, pembentukan fitur menggunakan TF-IDF Character N-Gram, serta penyeimbangan data menggunakan Random Undersampling. Pengukuran kemiripan menggunakan Cosine Distance, sedangkan sistem diimplementasikan dalam bentuk aplikasi web menggunakan Laravel. Hasil pengujian menunjukkan bahwa model terbaik diperoleh pada $K = 7$ dengan $K=7$ dengan accuracy 94.67%, Macro F1 94.54%, dan Top-5 Accuracy 98%. Hasil tersebut menunjukkan bahwa metode KNN mampu memberikan rekomendasi diagnosis penyakit dengan baik sehingga dapat digunakan sebagai sistem pendukung diagnosis awal

Kata Kunci: K-Nearest Neighbors, Klasifikasi Penyakit, Data Mining, TF-IDF, Cosine distance.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pelayanan kesehatan masyarakat merupakan salah satu dalam pembangunan nasional, khususnya dalam upaya meningkatkan derajat kesehatan masyarakat secara menyeluruh. Puskesmas sebagai fasilitas pelayanan kesehatan tingkat pertama memiliki peran strategis dalam memberikan pelayanan kesehatan dasar kepada masyarakat, termasuk pelayanan rawat jalan. Dalam pelaksanaannya, Puskesmas menghadapi tantangan yang kompleks, salah satunya adalah tingginya variasi penyakit yang diderita oleh pasien rawat jalan yang beragam. Kondisi ini menuntut adanya sistem pengelolaan data yang efektif dan efisien agar pelayanan kesehatan dapat diberikan secara optimal dan tepat sasaran. Puskesmas merupakan unit pelaksana teknis (UPT) Dinas Kesehatan yang bertanggung jawab menyelenggarakan pembangunan kesehatan di wilayah kerjanya, sebagai peyelenggara pembangunan kesehatan, puskesmas bertanggung jawab menyelenggarakan upaya kesehatan perorangan dan upaya kesehatan masyarakat. Secara umum, pengelolaan data pasien di Puskesmas masih dilakukan secara konvensional, sehingga seringkali menyulitkan dalam proses analisis data, khususnya dalam mengidentifikasi pola penyakit berdasarkan kunjungan pasien.

Puskesmas menghasilkan catatan pelayanan pasien dalam jumlah besar setiap harinya. Salah satu pemanfaatan data tersebut adalah membangun sistem yang mampu mengklasifikasikan penyakit berdasarkan gejala atau keluhan yang disampaikan pasien. Sistem semacam ini dapat menjadi alat bantu skrining awal serta mempercepat proses pencatatan diagnosa. Penelitian ini menggunakan algoritma K-Nearest Neighbors (KNN), yaitu metode klasifikasi berbasis kemiripan. Suatu data baru diklasifikasikan berdasarkan mayoritas kelas dari sejumlah k data terdekat. Karena gejala berupa teks, keluhan terlebih dahulu diubah menjadi representasi numerik menggunakan pembobotan TF-IDF (Term Frequency-Inverse Document Frequency) sebelum dihitung jaraknya. Dalam pelayanan kesehatan, klasifikasi penyakit pasien rawat jalan menjadi sangat relevan untuk dilakukan. Klasifikasi ini tidak hanya membantu tenaga medis dalam menentukan prioritas penanganan, tetapi juga memberikan gambaran yang jelas mengenai distribusi penyakit di wilayah kerja Puskesmas. Dengan demikian, Puskesmas dapat merancang intervensi kesehatan yang lebih tepat sasaran dan sesuai dengan kebutuhan masyarakat. Namun, proses klasifikasi penyakit secara manual memerlukan waktu dan tenaga yang tidak sedikit, serta rentan terhadap kesalahan manusia (*human error*). Oleh karena itu, diperlukan suatu metode yang mampu melakukan klasifikasi penyakit secara otomatis, cepat, dan akurat. Klasifikasi adalah teknik untuk merancang fungsi berdasarkan hasil pengamatan dari data dan atribut dari data tersebut sehingga dapat dilakukan pemetaan

terhadap data yang belum memiliki kelas kedalam data yang telah terklasifikasi sesuai dengan aturan-aturan yang diberikan. klasifikasi dapat dimanfaatkan untuk mengelompokkan penyakit pasien berdasarkan gejala dan atribut medis lainnya.

K-Nearest Neighbors (KNN) merupakan salah satu metode dalam *machine learning* yang termasuk ke dalam kategori supervised learning, yang digunakan untuk melakukan klasifikasi berdasarkan kedekatan jarak antara data uji dengan data latih. Metode KNN bekerja dengan cara mencari jumlah K data terdekat (nearest neighbors) dari data baru, kemudian menentukan kelas atau nilai dari data tersebut berdasarkan mayoritas kelas (untuk klasifikasi) atau rata-rata nilai (untuk regresi) dari data tetangga terdekat tersebut. Kedekatan antar data biasanya dihitung menggunakan metode pengukuran jarak.

1.2 Rumusan Masalah

Berdasarkan uraian diatas maka dirumuskan masalah yang akan diteliti yaitu berdasarkan uraian latar belakang diatas, penulis merumuskan masalah dalam penelitian adalah sebagai berikut:

1. Bagaimana penerapan metode *K-Nearest Neighbors* (KNN) dalam mengklasifikasikan penyakit pasien rawat jalan?
2. Bagaimana pengaruh nilai K terhadap klasifikasi penyakit pasien rawat jalan?
3. Bagaimana hasil klasifikasi penyakit pasien rawat jalan yang dihasilkan oleh metode *K-Nearest Neighbors* (KNN)?

1.3 Batasan Masalah

Agar dalam pembuatan proposal ini lebih terarah, maka penelitian ini berfokus pada pembahasan sebagai berikut:

1. Data yang digunakan adalah data pasien rawat jalan yang diperoleh dari pelayanan kesehatan.
2. Atribut yang digunakan meliputi data numeric dan kategorikal, seperti usia, jenis kelamin, suhu tubuh, dan gejala penyakit
3. Penyakit yang dikasifikasikan dibatasi pada beberapa jenis penyakit tertentu sesuai dengan data yang tersedia.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. mengetahui dan menerapkan metode *K-Nearest Neighbors* (KNN) dalam proses klasifikasi penyakit pasien rawat jalan berdasarkan data medis pasien.
2. menganalisis pengaruh variasi nilai K terhadap hasil klasifikasi penyakit pasien rawat jalan.
3. mengukur dan mengetahui tingkat akurasi metode KNN dalam mengklasifikasikan penyakit pasien rawat jalan.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Membantu Puskesmas dalam mengolah dan menganalisis data kunjungan pasien secara lebih efektif dan efisien.

2. Mempermudah tenaga kesehatan dalam mengidentifikasi pola penyakit yang berkembang di masyarakat.
3. Mendukung proses pengambilan keputusan, perencanaan program kesehatan, dan alokasi sumber daya berdasarkan data.

1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini memiliki 5 bab yaitu sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini membahas mengenai Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian dan sistematika penulisan.

BAB II TINJAUAN LITERATUR

Pada bab II terdapat 10 jurnal penelitian terdahulu yang mirip dengan penelitian yang penulis lakukan, untuk membandingkan dengan penelitian yang akan penulis lakukan, dan rangkuman dari hasil tinjauan penelitian tersebut.

BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang alur penelitian, peralatan dan bahan penelitian serta data yang digunakan saat penelitian.

1

BAB II

TINJAUAN LITERATUR

2.1 Pendahuluan

Berisikan tinjauan literatur yang membahas tentang penelitian yang telah dilakukan terdahulu oleh peneliti lain yang dijadikan perbandingan ataupun referensi peneliti untuk memperkuat penelitian yang dilakukan oleh peneliti sesuai dengan referensinya

2.2 tinjauan literature

Berikut penelitian yang telah dilakukan memiliki korelasi searah dengan penelitian yang akan dibahas dalam menyempurnakan perlu dilakukan kajianliteratur diantaranya;

Tabel 2. 1 Peneliti Terdahulu

No (1)	Nama Penulis (2)	Judul penelitian (3)	Pembahasan (4)
1	Alayda Zaielamy Putri & Muhammad Daud (2026)	Klasifikasi Penyakit Jantung Berbasis Data Rekam Medis Menggunakan Algoritma Local Mean K-Nearest Neighbor	Hasil penelitian menunjukkan bahwa algoritma Local Mean K-Nearest Neighbor (LMKNN) mampu mengklasifikasikan penyakit jantung berdasarkan data rekam medis dengan cukup baik karena dapat mengurangi pengaruh noise dan outlier melalui pendekatan rata-rata lokal. Hasil pengujian memperoleh akurasi 71,60%, dengan performa terbaik pada deteksi Penyakit Jantung Koroner yang memiliki tingkat recall 97,14%. Selain itu, faktor yang paling berpengaruh dalam klasifikasi adalah

3

3

			nyeri dada, diikuti oleh heart rate dan respiratory rate. Meskipun demikian, model masih mengalami kesulitan dalam mengklasifikasikan kelas dengan jumlah data yang sedikit, seperti gagal jantung, sehingga diperlukan perbaikan pada penelitian selanjutnya untuk meningkatkan kinerja model.
2	Ahmad Yogianto & Ahmad Homaidi (2024)	Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung	Hasil penelitian menunjukkan bahwa metode K-Nearest Neighbors (KNN) berhasil diterapkan untuk melakukan klasifikasi penyakit jantung menggunakan 303 data pasien dengan 14 atribut yang relevan. Dengan nilai K terbaik yaitu K=5, model menghasilkan akurasi sebesar 64,03%, presisi 64,58%, dan recall untuk pasien positif sebesar 75,15%. Secara keseluruhan, performa KNN dalam penelitian ini tergolong sedang dan masih dapat ditingkatkan dengan memperbaiki kualitas data atau menggabungkannya dengan algoritma lain seperti Decision Tree untuk memperoleh hasil klasifikasi yang lebih akurat
3	Dyan Dyanmita Putri & M.Tanzil Furqon (2020)	Klasifikasi Penyakit Kulit Pada Manusia Menggunakan Metode Binary Decision Tree Support Vector Machine (BDTSVM) (Studi Kasus: Puskesmas Dinoyo Kota Malan	Hasil penelitian menunjukkan bahwa implementasi metode hibrida BDTSVM sangat efektif dalam mengidentifikasi dan mengklasifikasikan jenis penyakit kulit. Model ini mencapai nilai akurasi terbaik sebesar {97,14%}. Disimpulkan bahwa bantuan teknologi komputer ini memiliki potensi besar untuk membantu diagnosis dini penyakit kulit.

Tabel 2.2 peneliti terdahulu (lanjutan)

(1)	(2)	(3)	(4)
4	Novita Ranti Muntari & Kharis Hudaiby Hanif (2022)	Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning	Penelitian ini membandingkan 7 algoritma Machine Learning, termasuk Decision Tree, untuk mengklasifikasikan jenis kanker payudara (ganas vs. jinak) menggunakan data Wisconsin Breast Cancer. Hasilnya, empat algoritma, yaitu Logistic Regression, Decision Tree, Naïve Bayes, dan K-Nearest Neighbor, menunjukkan nilai akurasi tertinggi yang sama, yaitu {95,00%}. Hal ini menyimpulkan bahwa algoritma-algoritma tersebut sangat efektif dalam memprediksi penentuan jenis kanker payudara.
5	Siti Kalimah (2022)	Klasifikasi Penyakit Diabetes Menggunakan Metode Decision Tree Dan Random Forest	Hasil penelitian menunjukkan bahwa kedua algoritma memberikan akurasi yang tinggi, di mana algoritma Decision Tree mencapai akurasi 79,00% dan Random Forest sedikit lebih unggul dengan akurasi 79,69%. Penelitian ini menyimpulkan bahwa faktor-faktor seperti Glucose, Body Mass Index (BMI), dan Age adalah variabel paling berpengaruh dalam klasifikasi penyakit diabetes.
6	Pinasti Ning Pragati (2024)	Klasifikasi Penyakit Terbanyak Di Rsud Ir. Soekarno Kabupaten Sukoharjo Menggunakan Decision Tree Algoritma C4.5	Hasil penelitian menunjukkan bahwa klasifikasi penyakit terbanyak menghasilkan 18 aturan (rule) dan mengidentifikasi blok-blok penyakit utama (seperti I00-I99, N00-N99, dan S00-T98). Namun, hasil evaluasi kinerja model klasifikasi penyakit terbanyak menggunakan confusion matrix menunjukkan akurasi yang sangat rendah, yaitu sebesar {23,48%}

Tabel 2.2 peneliti terdahulu (lanjutan)

(1)	(2)	(3)	(4)
7	Fida Maisa Hana (2020)	Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.	Hasil pengujian model Decision Tree C4.5 menunjukkan kinerja yang sangat baik dengan akurasi sebesar {97,12%}, Presisi sebesar {93,02%}, dan Recall mencapai {100,00%}. Penelitian ini menyimpulkan bahwa algoritma C4.5 efektif untuk meramalkan penyakit diabetes dan dapat membantu deteksi dini guna merendahkan komplikasi penyakit.
8	Siskawati Rahayu & Yuni Yamasari (2024)	Klasifikasi Penyakit Stroke dengan Metode Support Vector Machine (SVM)	Penelitian ini mengklasifikasi penyakit stroke menggunakan metode SVM dengan empat jenis kernel. Hasilnya, kernel polynomial memberikan performa terbaik dengan akurasi 78,86% , presisi 73,98% , dan recall 56,75% pada rasio 80:20. Kernel linear dan RBF menghasilkan kinerja cukup baik, sedangkan kernel sigmoid memiliki akurasi terendah. Kesimpulannya, kernel polynomial merupakan model paling optimal untuk klasifikasi stroke pada dataset ini.
9	Made Agus Oka Gunawan & Dewa Ayu Indah Saraswati (2023)	Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer	Hasil pengujian model klasifikasi menunjukkan tingkat Akurasi sebesar 80,43% dengan error classification 19,57%. Selain itu, diperoleh juga nilai AUC (Area Under Curve) sebesar 0,798. Penelitian menyimpulkan bahwa atribut yang paling menentukan (simpul akar atau root node dalam pohon keputusan) adalah atribut nominal "ST_Lereng" (ST_Slope) berdasarkan nilai gain ratio tertinggi

Tabel 2.2 peneliti terdahulu (lanjutan)

(1)	(2)	(3)	(4)
10.	Yuni Widiastiwi & Iin Ernawati (2021)	Klasifikasi Penyakit Batu Ginjal Menggunakan Algoritma Decision Tree C4.5 Dengan Membandingkan Hasil Uji Akurasi	Hasil penelitian menunjukkan bahwa pembagian data latih sebanyak 70% menghasilkan tingkat akurasi terbaik, yaitu sebesar 95,71%. Penelitian ini juga berhasil membentuk pohon keputusan dan menghasilkan aturan (rule) yang dapat mengidentifikasi faktor-faktor penyebab penyakit batu ginjal, di mana atribut seperti ureum, kalsium, dan keratin in menjadi penentu utama dalam klasifikasi.

Berdasarkan penelitian-penelitian diatas, maka penelitian ini akan membuat klasifikasi penyakit berdasarkan pola kunjungan pasien rawat jalan di puskesmas menggunakan metode decision tree yang akan membantu Puskesmas dalam mengklasifikasikan penyakit pasien rawat jalan berdasarkan pola kunjungan, sehingga dapat mendukung pengambilan keputusan dan peningkatan kualitas pelayanan kesehatan. Terdapat 10 jurnal yang diangkat sebagai perbandingan penelitian terdahulu, adapun perbandingan nya ialah:

Hasil penelitian menunjukkan bahwa algoritma Local Mean K-Nearest Neighbor (LMKNN) mampu mengklasifikasikan penyakit jantung berdasarkan data rekam medis dengan cukup baik karena dapat mengurangi pengaruh noise dan outlier

melalui pendekatan rata-rata lokal. Hasil pengujian memperoleh akurasi 71,60%, dengan performa terbaik pada deteksi Penyakit Jantung Koroner yang memiliki tingkat recall 97,14%. Selain itu, faktor yang paling berpengaruh dalam klasifikasi adalah nyeri dada, diikuti oleh heart rate dan respiratory rate. Meskipun demikian, model masih mengalami kesulitan dalam mengklasifikasikan kelas dengan jumlah data yang sedikit, seperti gagal jantung, sehingga diperlukan perbaikan pada penelitian selanjutnya untuk meningkatkan kinerja model.[1]

Penelitian ini menunjukkan bahwa metode *K-Nearest Neighbors* (KNN) dapat digunakan untuk mengklasifikasikan penyakit jantung menggunakan 303 data pasien dengan 14 atribut. Implementasi KNN dengan parameter $K = 5$ menghasilkan akurasi 64,03%, presisi 64,58%, dan recall 75,15%. Hasil ini menunjukkan bahwa metode KNN cukup efektif dalam membantu identifikasi pasien yang berpotensi menderita penyakit jantung, namun tingkat akurasi masih perlu ditingkatkan melalui perbaikan kualitas data atau kombinasi dengan metode klasifikasi lain.[2]

Penelitian ini membuktikan bahwa metode *Binary Decision Tree Support Vector Machine* (BDTSVM) efektif digunakan untuk mengklasifikasikan penyakit kulit pada manusia. Dengan menggunakan 150 data pasien dan 14 gejala, sistem mampu mengklasifikasikan 6 jenis penyakit kulit dengan akurasi terbaik sebesar 97,14%. Hasil penelitian menunjukkan bahwa pemilihan parameter SVM yang tepat sangat berpengaruh terhadap akurasi, sehingga metode BDTSVM berpotensi menjadi alat bantu diagnosis dini penyakit kulit yang cepat dan akurat.[3]

110 Penelitian ini membuktikan bahwa metode machine learning efektif
digunakan untuk mengklasifikasikan jenis kanker payudara (ganas dan jinak). Dari
39 tujuh algoritma yang diuji, yaitu logistic regression, decision tree, naïve bayes, dan k-
54 nearest neighbor menunjukkan performa terbaik dengan akurasi tertinggi sebesar
95%. Hasil ini menunjukkan bahwa algoritma-algoritma tersebut mampu membantu
43 pengambilan keputusan secara cepat dan akurat dalam penentuan jenis kanker
payudara, sehingga berpotensi mendukung diagnosis dini dan peningkatan kualitas
layanan kesehatan.[4]

43 Penelitian ini menunjukkan bahwa penerapan metode machine learning
mampu membantu proses klasifikasi dan pengambilan keputusan secara lebih efektif
28 dan akurat. Hasil pengujian membuktikan bahwa model yang digunakan memiliki
107 tingkat kinerja yang baik dalam mengolah data dan menghasilkan prediksi, sehingga
dapat dimanfaatkan sebagai alat bantu pendukung keputusan. Dengan demikian,
penerapan metode ini berpotensi meningkatkan efisiensi, ketepatan analisis, serta
kualitas hasil dalam bidang yang diteliti.[5]

123 Penerapan metode decision tree Algoritma C4.5 berhasil mengklasifikasikan
penyakit terbanyak di RSUD Ir. Soekarno dengan hasil yang sama antara Microsoft
Excel dan RapidMiner, yaitu beberapa kelompok penyakit utama. Namun, tingkat
akurasi model masih rendah sehingga diperlukan data yang lebih banyak untuk
meningkatkan hasil klasifikasi.[6]

31 Algoritma Decision Tree C4.5 efektif digunakan untuk mengklasifikasikan penderita penyakit diabetes berdasarkan 16 atribut gejala, dengan hasil kinerja yang sangat baik. Pengujian menghasilkan akurasi 97,12%, precision 93,02%, dan recall 100%, sehingga metode ini dinilai sangat akurat untuk deteksi dini penyakit diabetes.[7]

12 Metode Support Vector Machine (SVM) efektif digunakan untuk klasifikasi penyakit stroke dengan pengujian empat kernel. Hasil terbaik diperoleh pada kernel polynomial dengan rasio data 80:20, yang menghasilkan akurasi 78,86%, sehingga kernel polynomial menjadi model paling optimal dalam penelitian ini.[8]

104 Algoritma Decision Tree C4.5 berhasil mengklasifikasikan penyakit jantung dengan cukup baik menggunakan data Kaggle, menghasilkan akurasi 80,43% dan nilai AUC 0,798, sehingga layak digunakan sebagai metode pendukung keputusan diagnosis penyakit jantung.[9]

3 Algoritma Decision Tree C4.5 efektif digunakan untuk mengklasifikasikan penyakit batu ginjal berdasarkan data rekam medis pasien. Hasil pengujian menunjukkan akurasi terbaik sebesar 95,71% pada pembagian data latih 70% dan data uji 30%, sehingga metode ini mampu mengidentifikasi faktor penyebab batu ginjal dengan baik dan akurat.[10]

2.3 Rangkuman

89 Berdasarkan rangkuman penelitian terdahulu, dapat disimpulkan bahwa
54 penerapan metode machine learning dalam bidang kesehatan telah menunjukkan hasil
yang sangat positif, khususnya dalam proses klasifikasi penyakit dan pendukung
pengambilan keputusan medis K-Nearest Neighbors (KNN) menjadi metode yang
paling banyak digunakan dan terbukti efektif dalam mengklasifikasikan berbagai
jenis penyakit, seperti diabetes, penyakit jantung, batu ginjal, serta diagnosis pasien
di klinik dan rumah sakit, dengan tingkat akurasi yang relatif tinggi, bahkan mencapai
30 lebih dari 95% pada beberapa penelitian. Keunggulan utama metode ini terletak pada
kemampuannya dalam menghasilkan aturan (rule) yang mudah dipahami serta
mengidentifikasi atribut-atribut penting yang berpengaruh terhadap kondisi pasien.

36 Selain K-Nearest Neighbors (KNN), metode lain seperti Decision Tree dan
3 Support Vector Machine (SVM) juga menunjukkan performa yang cukup baik dalam
klasifikasi penyakit tertentu, seperti penyakit jantung, penyakit kulit, dan stroke.
Namun, beberapa penelitian menunjukkan bahwa akurasi metode tersebut masih
26 sangat dipengaruhi oleh kualitas data, jumlah atribut, serta pemilihan parameter dan
130 rasio data latih dan data uji.

Secara keseluruhan, penelitian-penelitian terdahulu membuktikan bahwa
penggunaan algoritma machine learning mampu meningkatkan kecepatan, ketepatan,
dan akurasi dalam proses klasifikasi penyakit. Implementasi metode ini tidak hanya
membantu tenaga medis dalam pengambilan keputusan, tetapi juga berkontribusi

19 pada peningkatan kualitas pelayanan kesehatan dan efisiensi operasional fasilitas kesehatan. Oleh karena itu, penelitian mengenai klasifikasi penyakit pasien rawat jalan di puskesmas menggunakan metode *K-Nearest Neighbors* (KNN) sangat relevan untuk dikembangkan sebagai solusi pendukung dalam meningkatkan kualitas layanan kesehatan.

BAB III

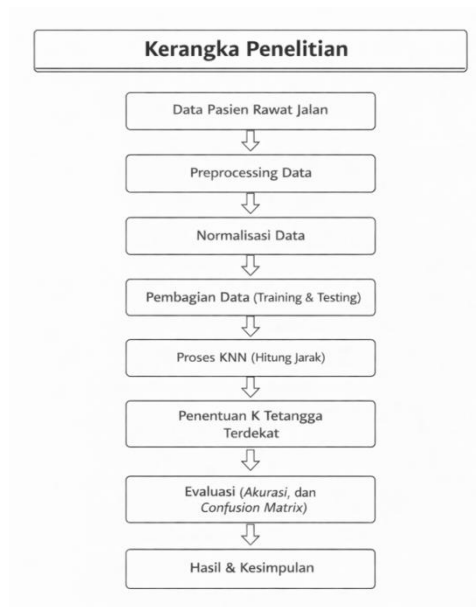
METODE PENELITIAN

3.1 Pendahuluan

Bab ini menjelaskan langkah-langkah penelitian yang dilakukan untuk mengklasifikasi penyakit pasien rawat jalan di Puskesmas menggunakan metode *K-Nearest Neighbors* (KNN). Metodologi penelitian mencakup alur penelitian, jenis dan sumber data, teknik pengumpulan data, proses pengolahan data, metode analisis, serta alat yang digunakan dalam penelitian. Penelitian ini menggunakan metode kuantitatif dengan pendekatan data mining. Teknik yang digunakan adalah klasifikasi, dengan algoritma *K-Nearest Neighbors* (KNN) untuk memodelkan pola kunjungan pasien berdasarkan atribut-atribut tertentu sehingga dapat memprediksi jenis penyakit rawat jalan.

3.2 kerangka penelitian

Kerangka penelitian dirancang memberikan panduan yang sistematis dan terstruktur dalam pelaksanaan penelitian, Berikut adalah kerangka dalam penelitian ini :



Gambar 1 Kerangka Penelitian

3.3 Data Pasien Rawat Jalan

Pengumpulan data dilakukan untuk memperoleh informasi mendalam terkait kebutuhan jenis data yang diperlukan yaitu Mengambil data historis kunjungan pasien rawat jalan dari Puskesmas, termasuk atribut seperti umur, jenis kelamin, waktu kunjungan, poli tujuan, dan diagnosa.

3.4 Preprocessing Data

Preprocessing data merupakan tahap awal dalam penelitian yang bertujuan untuk menyiapkan data pasien rawat jalan agar siap digunakan dalam proses klasifikasi menggunakan metode *K-Nearest Neighbors* (KNN). Tahap ini dilakukan untuk meningkatkan kualitas data sehingga hasil klasifikasi menjadi lebih akurat[11]

139 B. Proses preprocessing meliputi data cleaning, yaitu pembersihan data dari nilai yang hilang, data duplikat, dan kesalahan penulisan; data transformation, yaitu mengubah data kategorikal menjadi bentuk numerik; serta data normalization, yaitu menyamakan skala data agar perhitungan jarak pada metode KNN dapat dilakukan secara optimal.

91 Tahap data cleaning dilakukan dengan membersihkan data dari nilai yang hilang (missing value), data yang tidak konsisten, serta data yang duplikat[12]. Data yang tidak lengkap dapat dihapus atau diperbaiki agar tidak mempengaruhi hasil analisis. 16 Selain itu, kesalahan penulisan pada data pasien juga diperbaiki untuk menjaga keakuratan data.

83 Tahap data transformation, yaitu mengubah data ke dalam bentuk yang sesuai untuk proses klasifikasi. Misalnya, data kategorikal seperti jenis kelamin atau gejala diubah menjadi bentuk numerik agar dapat diproses oleh metode KNN. Contohnya, 4 “Laki-laki” diubah menjadi 1 dan “Perempuan” menjadi 0, atau gejala “Ya” menjadi 1 dan “Tidak” menjadi 0.

129 Tahap data normalization, yaitu proses penyetaraan skala data numerik agar berada dalam rentang yang sama. Hal ini penting karena metode KNN menggunakan 4 perhitungan jarak, sehingga perbedaan skala antar atribut dapat mempengaruhi hasil klasifikasi[13]. 24 Normalisasi dapat dilakukan menggunakan metode seperti Min-Max Normalization.

3.5 Normalisasi Data

Normalisasi data merupakan proses penyesuaian skala nilai pada atribut numerik agar berada dalam rentang yang sama. Pada penelitian dengan judul “Klasifikasi Penyakit Pasien Rawat Jalan Menggunakan Metode K-Nearest Neighbors (KNN)”, normalisasi sangat penting karena metode KNN menggunakan perhitungan jarak antar data, sehingga perbedaan skala antar atribut dapat mempengaruhi hasil klasifikasi[14]. Atribut seperti usia, suhu tubuh, dan tekanan darah memiliki rentang nilai yang berbeda, sehingga perlu dinormalisasi agar tidak mendominasi perhitungan jarak. Metode normalisasi yang digunakan dalam penelitian ini adalah Min-Max Normalization, dengan rumus:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

di mana:

X = nilai data asli

$X_{\min}$ = nilai minimum

$X_{\max}$ = nilai maksimum

X' = nilai hasil normalisasi

Dengan dilakukan normalisasi, setiap atribut memiliki kontribusi yang seimbang dalam perhitungan jarak, sehingga dapat meningkatkan akurasi hasil klasifikasi menggunakan metode KNN.

3.6 Pembagian Data

Pembagian data merupakan tahap dalam penelitian yang bertujuan untuk memisahkan dataset menjadi data latih (training data) dan data uji (testing data)[15].

Pada penelitian dengan judul “Klasifikasi Penyakit Pasien Rawat Jalan Menggunakan Metode K-Nearest Neighbors (KNN)”, pembagian data dilakukan untuk menguji kemampuan metode KNN dalam mengklasifikasikan data pasien secara akurat.

Data dibagi secara stratified menjadi tiga bagian agar proporsi tiap penyakit terjaga: 70% train, 15% test. Pembagian ini penting agar evaluasi bersifat objektif data test sama sekali tidak dilibatkan pada tahap pemilihan parameter. Peran masing-masing subset dirangkum pada tabel 3.1

Tabel 3.1 Pembagian data dan fungsinya

Subset	Jumlah data	Fungsi
Train	700	Melatih model KNN
Validation	150	Memilih nilai k terbaik
Test	150	Evaluasi akhir yang objektif

3.7 Proses KNN

Proses K-Nearest Neighbors (KNN) merupakan tahap utama dalam penelitian ini, yaitu menghitung jarak antara data uji dengan data latih untuk menentukan kedekatan antar data. Pada penelitian dengan judul “Klasifikasi Penyakit Pasien Rawat Jalan

Menggunakan Metode K-Nearest Neighbors (KNN)”, perhitungan jarak digunakan untuk mengetahui tingkat kemiripan kondisi pasien berdasarkan atribut yang dimiliki.

Metode yang digunakan untuk menghitung jarak adalah Euclidean Distance, yaitu metode yang mengukur jarak antara dua titik dalam ruang multidimensi. Rumus Euclidean Distance adalah sebagai berikut:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

di mana:

d = jarak antara data uji dan data latih

x_{ix_ixi} = nilai atribut ke- i pada data uji

y_{iy_iyi} = nilai atribut ke- i pada data latih

nnn = jumlah atribut

Setelah seluruh jarak dihitung, langkah selanjutnya adalah mengurutkan jarak dari yang terkecil hingga terbesar untuk menentukan K tetangga terdekat. Data dengan jarak paling kecil dianggap memiliki kemiripan tertinggi dan akan digunakan dalam proses penentuan kelas penyakit pasien.

Dengan menggunakan perhitungan jarak ini, metode KNN dapat mengklasifikasikan data pasien baru berdasarkan kemiripan dengan data pasien sebelumnya.

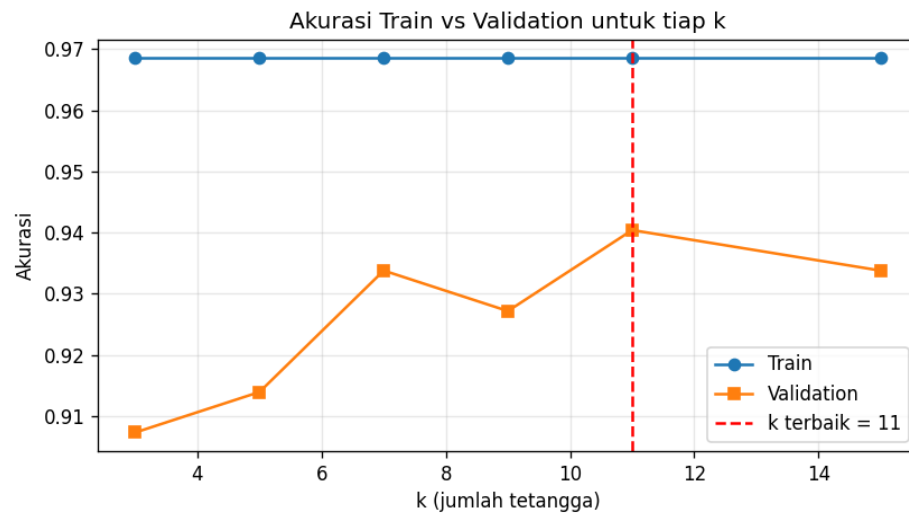
3.8 Penentuan K Tetangga Terdekat

Penentuan K tetangga terdekat merupakan tahap lanjutan dalam metode K-Nearest Neighbors (KNN) setelah proses perhitungan jarak dilakukan. Pada penelitian dengan judul “Klasifikasi Penyakit Pasien Rawat Jalan Menggunakan Metode K-Nearest Neighbors (KNN)”, nilai K menunjukkan jumlah data tetangga terdekat yang digunakan sebagai acuan dalam menentukan kelas suatu data uji.

Nilai K terbaik ditentukan berdasarkan akurasi tertinggi pada data validation, dengan kandidat K=3, 5, 9, 11, dan 15. Akurasi pada data train data juga dipantau untuk mendeteksi overviting. hasilnya dapat dilihat pada tabel 3.2 dan gambar 3.1

Tabel 3.2 Akurasi train dan validation untuk tiap nilai k

Nilai k	Akurasi Train	Akurasi Validation	Keterangan
3	0,969	0,907	
5	0,969	0,914	
7	0,969	0,934	
9	0,969	0,927	
11	0,969	0,940	Terbaik
15	0,969	0,934	



Gambar 3 Kurva akurasi train vs validation terhadap nilai k

Berdasarkan Tabel 3.2, nilai K=11 memberikan akurasi validation tertinggi (0,940) dan dipilih sebagai parameter model akhir. Setelah K terpilih, model final dilatih ulang menggunakan gabungan data train+validation, lalu diuji pada data test

3.9 Evaluasi Model

Evaluasi model merupakan tahap untuk mengukur kinerja metode K-Nearest Neighbors (KNN) dalam mengklasifikasikan penyakit pasien rawat jalan. Pada penelitian dengan judul “Klasifikasi Penyakit Pasien Rawat Jalan Menggunakan Metode K-Nearest Neighbors (KNN)”, evaluasi dilakukan menggunakan akurasi dan confusion matrix.

Akurasi digunakan untuk mengetahui tingkat ketepatan model dalam mengklasifikasikan data. Nilai akurasi diperoleh dari perbandingan antara jumlah data yang diklasifikasikan dengan benar terhadap seluruh data yang diuji. Rumus akurasi adalah sebagai berikut:

$$Akurasi = \frac{Jumlah\ data\ benar}{Total\ data\ uji} \times 100\%$$

Semakin tinggi nilai akurasi, maka semakin baik kinerja model dalam melakukan klasifikasi.

Selain itu, digunakan juga confusion matrix, yaitu tabel yang digunakan untuk mengevaluasi hasil klasifikasi dengan membandingkan antara nilai aktual dan hasil prediksi. Confusion matrix terdiri dari empat komponen, yaitu:

True Positive (TP): data yang diprediksi benar sesuai kelas positif

True Negative (TN): data yang diprediksi benar sesuai kelas negatif

False Positive (FP): data yang diprediksi positif tetapi sebenarnya negatif

False Negative (FN): data yang diprediksi negatif tetapi sebenarnya positif

Melalui confusion matrix, dapat diketahui performa model secara lebih rinci, tidak hanya dari sisi akurasi tetapi juga kesalahan klasifikasi yang terjadi.

Dengan menggunakan akurasi dan confusion matrix, diharapkan dapat diperoleh gambaran yang jelas mengenai tingkat keberhasilan metode KNN dalam klasifikasi penyakit pasien rawat jalan.

Hasil & kesimpulan

Pada tahap hasil, penelitian ini menyajikan output dari penerapan metode K-Nearest Neighbors (KNN) dalam mengklasifikasikan penyakit pasien rawat jalan.

Hasil yang diperoleh berupa kelas penyakit dari data uji, yang ditentukan berdasarkan

kemiripan dengan data latih menggunakan perhitungan jarak. Selain itu, ditampilkan juga hasil evaluasi model menggunakan akurasi dan confusion matrix untuk mengetahui tingkat kinerja metode KNN. Dari hasil pengujian, dapat diketahui seberapa baik metode KNN dalam mengklasifikasikan penyakit berdasarkan data pasien yang digunakan.

47 Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa
120 metode K-Nearest Neighbors (KNN) mampu digunakan untuk mengklasifikasikan
penyakit pasien rawat jalan dengan tingkat akurasi tertentu, tergantung pada nilai K
yang digunakan dan kualitas data. Metode ini bekerja dengan baik dalam
mengidentifikasi kemiripan antar data pasien, sehingga dapat membantu dalam proses
28 klasifikasi penyakit. Dengan demikian, metode KNN dapat dijadikan sebagai sistem
pendukung keputusan dalam bidang kesehatan, khususnya dalam membantu tenaga
medis dalam mengelompokkan penyakit pasien secara lebih cepat dan sistematis.

60

BAB IV

HASIL DAN PEMBAHASAN

65

Pada bab ini dijelaskan hasil dari penelitian yang telah dilakukan mengenai klasifikasi penyakit pasien rawat jalan menggunakan metode K-Nearest Neighbors (KNN). Sistem DiagnoKNN merupakan aplikasi klasifikasi diagnosis penyakit berbasis teks keluhan pasien. Sistem menerima input berupa keluhan atau gejala pasien, pilihan jenis kelamin, dan lama sakit. Model kemudian menghasilkan rekomendasi diagnosis dalam bentuk daftar Top-5 kelas penyakit dengan skor kecocokan.

4.1 Perbandingan Akurasi Antar Subset

Untuk memeriksa kemampuan generalisasi model, akurasi dibandingkan antara data train+validation dan data test (tabel 4.1). selisih yang kecil menandakan model tidak mengalami overfitting yang berarti.

Tabel 4.1 Perbandingan akurasi model final antar subset

Subset Data	Akurasi
Train+Validation	0,967
Test	0,920

.

4.2 Hasil Evaluasi lengkap pada Data Test

Model KNN dengan $K=11$ dievaluasi pada data test. Metrik lengkap disajikan pada tabel 4.2

Tabel 4.2 Metrix evaluasi lengkap pada data Test

Metrik	Nilai
Akurasi	0,920 (92,0%)
Precision (macro)	0,927
Precision (weighted)	0,927
Recall (macro)	0,920
Recall (weighted)	0,920
F1-score (macro)	0,921
F1-score (weighted)	0,921
Cohen's Kappa	0,911

Model ini mencapai akurasi 92.0% dengan Cohen's Kappa 0,911. Nilai Kappa diatas 0,8 menunjukkan tingkat kesepakatan antara prediksi dan label sebenarnya tergolong sangat baik (almost perfect agreement), bukan sekedar hasil kebetulan.

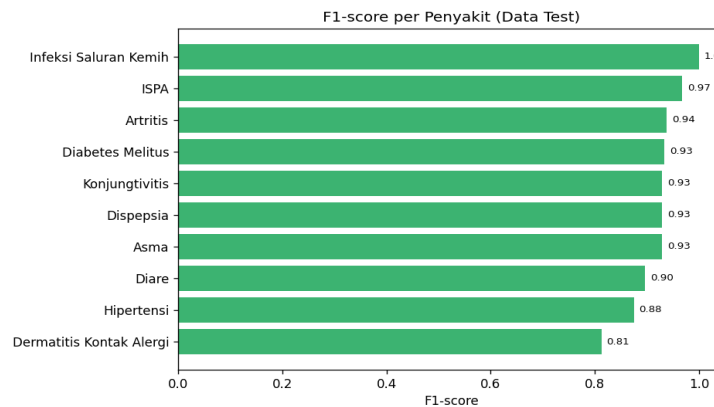
4.3 Performa per Penyakit

Rincian precisio, recall, dan F1-score tiap penyakit disajikan pada tabel 4.3 dan gambar 4.1

Tabel 4.3 *Classification report per penyakit (data test)*

Penyakit	Precision	Recall	F1-score	Data uji
Infeksi Saluran Kemih	1,00	1,00	1,00	15
ISPA	0,94	1,00	0,97	15
Artritis	0,88	1,00	0,94	15
Asma	1,00	0,87	0,93	15
Diabetes Melitus	0,93	0,93	0,93	15
Dispepsia	1,00	0,87	0,93	15
Konjungtivitis	1,00	0,87	0,93	15
Diare	0,93	0,87	0,90	15
Hipertensi	0,82	0,93	0,88	15
Dermatitis Kontak Alergi	0,76	0,87	0,81	15

Setelah proses preprosesing dan pembagian data selesai dilakukan, model KNN dilatih menggunakan data training dan uji menggunakan data testing.



Gambar 4.1 F1-score per penyakit pada data test

4.4 Fitur dan Target Klasifikasi

Fitur utama sistem dipilih agar model belajar dari gejala pasien, bukan dari konteks administrasi pelayanan. Kolom seperti poli ruangan, asuransi, tanggal kunjungan, dan identitas layanan tidak digunakan sebagai fitur utama untuk menghindari data leakage dan shortcut administratif.

Tabel 4.4. Fitur dan target yang digunakan dalam penelitian

Komponen	Keterangan
gejala_text	Teks keluhan atau gejala pasien yang telah dinormalisasi.
lama_sakit_bucket	Kategori lama sakit: 0 hari, 1-3 hari, 4-7 hari, 8-14 hari, dan >14 hari.
jenis_kelamin_code	Tersedia pada dataset dan form input, tetapi tidak digunakan pada model final karena eksperimen menunjukkan penurunan metrik validasi.
target_group_code	Kode kelas penyakit hasil pengelompokan diagnosis.
target_group_name	Nama kelas penyakit yang menjadi output klasifikasi.

Penggunaan fitur gejala_text menjadi komponen utama karena metode klasifikasi berbasis kemiripan teks sangat dipengaruhi oleh kualitas representasi gejala pasien.

4.5 Kelas Penyakit

Dataset terdiri atas 1.00 baris data dengan 10 kelas penyakit yang seimbang (masing-masing 100 baris). Daftar penyakit beserta kode ICD-10 ditunjukkan pada tabel 4.5

Tabel 4.5. Daftar kelas penyakit, kode ICD-10, dan jumlah data

No	Penyakit	Kode ICD	Jumlah	Label ID
1	ISPA (Infeksi Saluran Pernapasan Akut)	J06.9	100	7
2	Hipertensi	I10	100	6
3	Dispepsia	K30	100	5
4	Diabetes Melitus	E11	100	3
5	Arthritis (Radang Sendi)	M13.9	100	0
6	Dermatitis Kontak Alergi	L23	100	2
7	Asma	J45	100	1
8	Diare / Gastroenteritis	A09	100	4
9	Konjungtivitis	H10	100	9
10	Infeksi Saluran Kemih (ISK)	N39.0	100	8

4.6 Tahapan Praprocessing

Preprocessing dilakukan untuk mengubah data pelayanan pasien menjadi data teks yang dapat diproses oleh algoritma machine learning. Tahap ini penting karena kualitas teks keluhan sangat memengaruhi hasil klasifikasi.

1. Mengambil data mentah dari file laporan harian pelayanan pasien.
2. Membersihkan baris yang tidak memiliki diagnosis atau keluhan yang memadai.
3. Menormalisasi teks gejala, termasuk huruf kecil, pembersihan simbol, dan penggantian istilah tidak baku.
4. Mengelompokkan diagnosis asli ke dalam target_group_code dan target_group_name.
5. Membentuk fitur lama sakit ke dalam kategori bucket agar lebih konsisten.
6. Menghapus fitur administrasi seperti poli ruangan dan asuransi dari dataset training utama.
7. Menyimpan dataset final, audit dataset, mapping kelas, dan runtime model.

Tabel 4.6. Contoh normalisasi istilah pada teks gejala

Komponen	Keterangan
filek, flue, flu	dinormalisasi menjadi pilek
batpil	dinormalisasi menjadi batuk pilek
hifertensi	dinormalisasi menjadi hipertensi
tensi	dinormalisasi menjadi tekanan darah
gdp	dinormalisasi menjadi gula darah puasa

Komponen	Keterangan
gds	dinormalisasi menjadi gula darah sewaktu
tdk	dinormalisasi menjadi tidak
n2	dinormalisasi menjadi nyeri

4.7 Metode Klasifikasi

Metode yang digunakan adalah K-Nearest Neighbors (KNN). Algoritma ini bekerja dengan mencari data training yang paling mirip dengan input baru. Dalam sistem ini, kemiripan dihitung dari representasi teks gejala menggunakan TF-IDF berbasis karakter.

Tabel 4.7. Konfigurasi model klasifikasi

Komponen	Keterangan
Algoritma	K-Nearest Neighbors (KNN)
Nilai K terbaik	21
Metode jarak	Cosine distance
Pembobotan tetangga	Distance weighting, tetangga yang lebih dekat memiliki pengaruh lebih besar.
Vectorizer	Custom char_wb TF-IDF ngram_range=(3, 5) max_features=4000 sublinear_tf=True
Fitur model final	gejala_text, lama_sakit_bucket
Feature mode	symptom_duration

4.8 Penanganan Ketidakseimbangan Data

Distribusi awal dataset menunjukkan dominasi kelas mayoritas, terutama kelompok ISPA/flu/batuk-demam. Jika tidak ditangani, model cenderung lebih sering memilih kelas mayoritas walaupun input tidak sepenuhnya sesuai. Oleh karena itu, runtime final menggunakan random undersampling dengan batas maksimum 150 data per kelas.

Tabel 4.8. Strategi balancing data

Komponen	Keterangan
Metode	random_undersampling
Target maksimum	150 data per kelas mayoritas
Data runtime setelah resampling	1165 data
Alasan penggunaan	Mengurangi dominasi kelas mayoritas tanpa menambah data sintetis.

4.9 Evaluasi Model

Evaluasi dilakukan dengan membagi dataset menjadi data latih dan data uji. Jumlah data latih adalah 1458 baris dan data uji adalah 363 baris. Beberapa nilai K diuji untuk memperoleh konfigurasi yang paling baik berdasarkan accuracy, Top-3 accuracy, Top-5 accuracy, dan macro F1.

Tabel 4.9. Hasil evaluasi pada beberapa nilai K

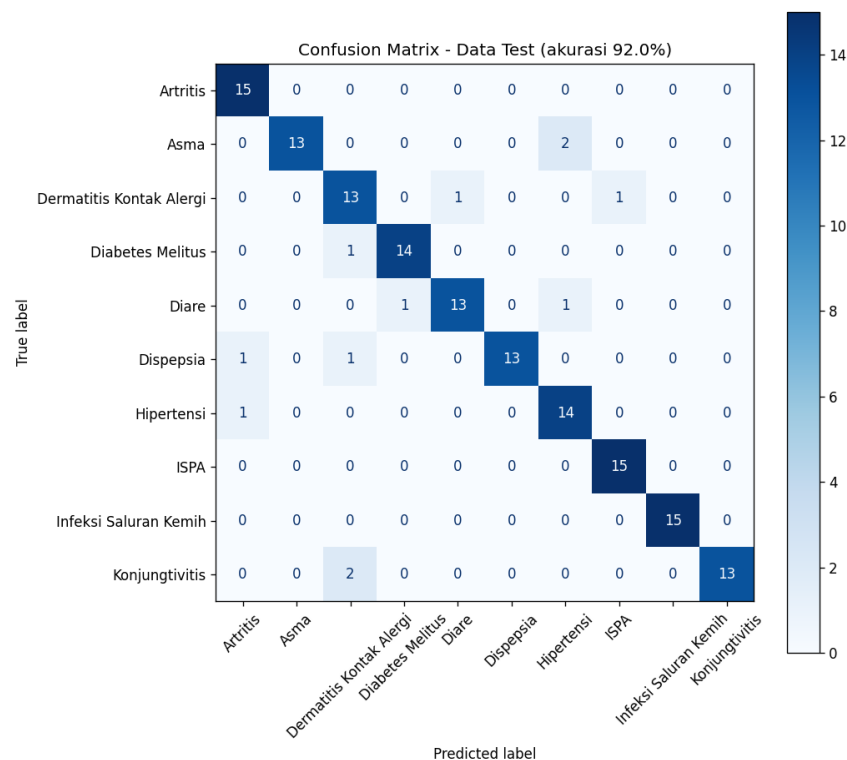
K	Val Accuracy	Macro F1	Weighted F1	Top-3	Top-5
3	94%	93.86%	93.86%	95.33%	98%
5	94.67%	94.53%	94.53%	95.33%	98%
7	94.67%	94.54%	94.54%	95.33%	98%
9	94.67%	94.67%	94.51%	95.33%	96.67%
11	94.67%	94.47%	94.47%	94.67%	94.67%
13	94.67%	94.47%	94.47%	94.67%	94.67%
15	94.67%	94.5%	94.5%	94.67%	96%
17	94.67%	94.47%	94.47%	95.33%	96.67%
19	94.67%	94.47%	94.47%	95.33%	96.67%
21	94.67%	94.47%	94.47%	94.67%	97.33%

Model terbaik diperoleh pada K = 21 dengan accuracy 94.67%, macro F1 94.47%, weighted F1 94.47%, Top-3 accuracy 94.67%, dan Top-5 accuracy 97.33%.

Nilai Top-3 dan Top-5 yang tinggi menunjukkan bahwa walaupun prediksi peringkat pertama belum selalu benar, diagnosis yang sesuai sering masih muncul pada daftar rekomendasi teratas. Hal ini relevan karena sistem dirancang sebagai rekomendasi diagnosis awal, bukan pengganti keputusan medis final.

Counfusion Matrix

Counfusion matrix pada gambar 4.2 memperlihatkan sebaran prediksi benar (diagonal) dan kesalahan klasifikasi antar penyakit pada data test.



Gambar 4 Confusion matrix hasil klasifikasi KNN pada data test (akurasi 92,0%)

Gambar tersebut menampilkan Confusion Matrix hasil evaluasi model K-Nearest Neighbors (KNN) pada data testing dengan tingkat akurasi sebesar 92,0%. Confusion Matrix digunakan untuk mengukur kemampuan model dalam mengklasifikasikan setiap kelas penyakit dengan membandingkan label sebenarnya (True Label) pada sumbu vertikal dengan label hasil prediksi (Predicted Label) pada sumbu horizontal. Setiap kotak pada matriks menunjukkan jumlah data yang diprediksi ke dalam kelas tertentu, sedangkan intensitas warna biru menunjukkan banyaknya jumlah data pada setiap sel. Nilai yang berada pada diagonal utama

merupakan prediksi yang benar (True Positive), sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi (misclassification). Dari hasil tersebut terlihat bahwa sebagian besar data berhasil diprediksi dengan benar, ditunjukkan oleh nilai diagonal yang tinggi, seperti Arthritis (15 data), Asma (13 data), Dermatitis Kontak Alergi (13 data), Diabetes Melitus (14 data), Diare (13 data), Dispepsia (13 data), Hipertensi (14 data), ISPA (15 data), Infeksi Saluran Kemih (15 data), dan Konjungtivitis (13 data). Hal ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali sebagian besar kelas penyakit.

Meskipun demikian, masih terdapat beberapa kesalahan klasifikasi yang ditunjukkan oleh nilai di luar diagonal. Misalnya, 2 data Asma diprediksi sebagai Hipertensi, 1 data Dermatitis Kontak Alergi diprediksi sebagai Diare dan 1 data sebagai ISPA, 1 data Diabetes Melitus diprediksi sebagai Dermatitis Kontak Alergi, 1 data Diare diprediksi sebagai Diabetes Melitus dan 1 data sebagai Hipertensi, 1 data Dispepsia diprediksi sebagai Arthritis dan 1 data sebagai Dermatitis Kontak Alergi, 1 data Hipertensi diprediksi sebagai Arthritis, serta 2 data Konjungtivitis diprediksi sebagai Dermatitis Kontak Alergi. Kesalahan-kesalahan tersebut menunjukkan bahwa beberapa penyakit memiliki karakteristik gejala yang mirip sehingga model mengalami kesulitan dalam membedakannya secara sempurna. Namun, karena jumlah kesalahan relatif sedikit dibandingkan jumlah prediksi yang benar, model tetap mampu mencapai akurasi sebesar 92,0%, sehingga dapat disimpulkan bahwa

53 algoritma KNN memiliki performa yang baik dan layak digunakan sebagai sistem pendukung klasifikasi diagnosis penyakit berdasarkan gejala pasien.

Grafik Performa Model Per Nilai K

97 Grafik Performa Model per Nilai K digunakan untuk memvisualisasikan hasil pengujian beberapa nilai K pada metode K-Nearest Neighbors (KNN). Grafik ini membantu peneliti dalam menentukan nilai K yang paling optimal berdasarkan hasil evaluasi model. Melalui grafik tersebut dapat diketahui perubahan performa model, baik peningkatan maupun penurunan akurasi, ketika nilai K diubah. Dengan adanya visualisasi grafik, proses pemilihan nilai K terbaik menjadi lebih mudah dibandingkan hanya melihat data dalam bentuk tabel.



Gambar 5 Grafiik Performa Model Per Nilai K

Berdasarkan grafik Performa Validasi per Nilai K pada Dashboard Admin DiagnokNN yang digunakan untuk mengevaluasi kinerja algoritma K-Nearest Neighbors (KNN) berdasarkan beberapa nilai K yang diuji. Grafik ini bertujuan untuk

121

90

menentukan nilai K yang paling optimal sebelum model digunakan pada proses prediksi. Sumbu horizontal menggambarkan variasi nilai K yang diuji, sedangkan sumbu vertikal menunjukkan tingkat performa model berdasarkan hasil validasi. Dari grafik terlihat bahwa performa model mengalami peningkatan pada nilai K awal, kemudian cenderung stabil pada nilai K berikutnya. Titik berwarna oranye menunjukkan bahwa $K = 7$ dipilih sebagai nilai terbaik karena memberikan hasil evaluasi yang paling optimal dan konsisten dibandingkan nilai K lainnya. Di sisi kanan grafik juga ditampilkan informasi " $K = 7 \mid \text{Test } 93,33\%$ ", yang menunjukkan bahwa model dengan nilai K sebesar 7 memperoleh akurasi pengujian sebesar 93,33% pada data testing. Selain digunakan untuk menentukan nilai K terbaik, grafik ini juga menunjukkan bahwa proses pemilihan parameter dilakukan menggunakan data validasi, sedangkan data testing hanya digunakan untuk mengukur performa akhir model. Dengan demikian, grafik ini membuktikan bahwa model KNN memiliki performa yang stabil pada berbagai nilai K dan bahwa pemilihan $K = 7$ mampu menghasilkan tingkat akurasi yang tinggi, sehingga model layak digunakan sebagai sistem pendukung klasifikasi diagnosis penyakit berdasarkan gejala pasien.

4.9 Implementasi Sistem

Aplikasi dibangun sebagai sistem web dengan backend Laravel dan frontend Vue/Inertia. Runtime model disimpan dalam format JSON sehingga proses prediksi dapat dijalankan langsung di PHP tanpa memanggil Python saat aplikasi digunakan.

Tabel 4.12. Komponen implementasi sistem

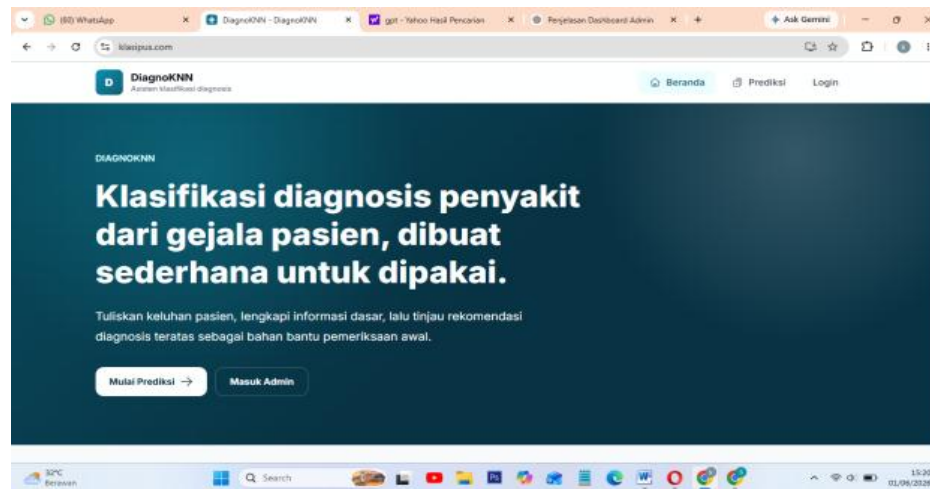
Komponen	Keterangan
Backend	Laravel 10
Frontend	Vue 3, Inertia.js, Tailwind CSS
Training model	Python script train_diagnosis_knn_v2.py
Runtime model	storage/app/knn/diagnosis_knn_runtime.json
Service prediksi	app/Services/DiagnosisPredictionService.php
Halaman prediksi	resources/js/Pages/Public/Predict.vue

4.10. Tampilan Aplikasi

Untuk memberikan visualisasi operasional komprehensif serta menyoroti kemudahan interaksi yang ditawarkan kepada pengguna, disajikan serangkaian tangkapan layar antarmuka website. Tampilan-tampilan ini secara sistematis mengilustrasikan alur interaksi pengguna dengan sistem, mulai dari fase inisiasi website hingga proses klasifikasi penyakit pasien rawat jalan

a. Tampilan awal

Tampilan awal website menyajikan pembuka yang sederhana dan menarik. Berikut tampilannya.



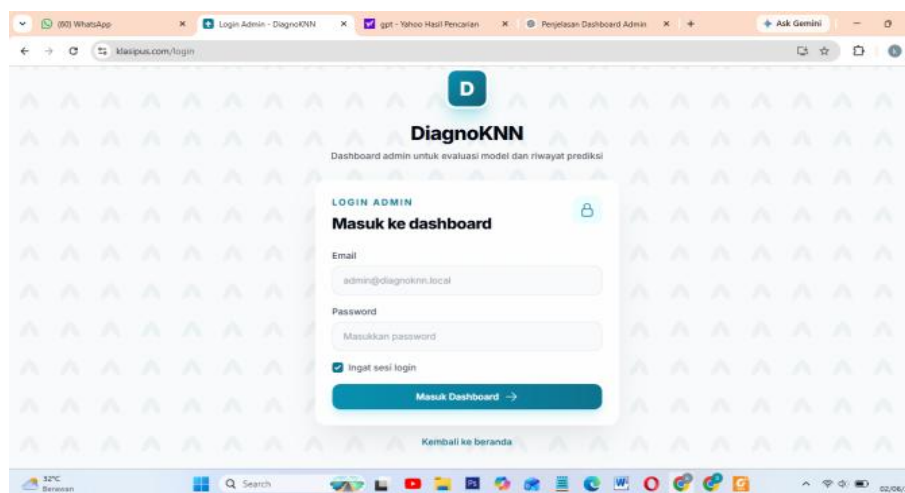
Gambar 6 Tampilan Website

Gambar tersebut menampilkan halaman beranda aplikasi DiagnokNN, yaitu sistem berbasis web yang digunakan untuk membantu klasifikasi diagnosis penyakit berdasarkan gejala pasien menggunakan algoritma K-Nearest Neighbor (KNN). Pada bagian atas terdapat logo aplikasi serta menu navigasi Beranda, Prediksi, dan Login yang memudahkan pengguna mengakses fitur sistem. Bagian utama halaman menampilkan judul “Klasifikasi diagnosis penyakit dari gejala pasien, dibuat sederhana untuk dipakai”, yang menjelaskan tujuan aplikasi sebagai alat bantu diagnosis awal yang mudah digunakan. Di bawahnya terdapat deskripsi singkat bahwa pengguna cukup memasukkan gejala dan informasi dasar pasien untuk memperoleh rekomendasi diagnosis. Selain itu, tersedia tombol Mulai Prediksi untuk mengakses fitur klasifikasi penyakit dan tombol Masuk Admin untuk mengakses halaman pengelolaan sistem. Secara keseluruhan, tampilan

halaman ini dirancang sederhana, informatif, dan mudah digunakan sehingga pengguna dapat memahami fungsi aplikasi serta langsung mengakses fitur utama yang tersedia.

b. Login admin

Halaman Login Admin pada aplikasi DiagnokNN yang berfungsi sebagai gerbang akses bagi administrator untuk masuk ke dashboard pengelolaan sistem.

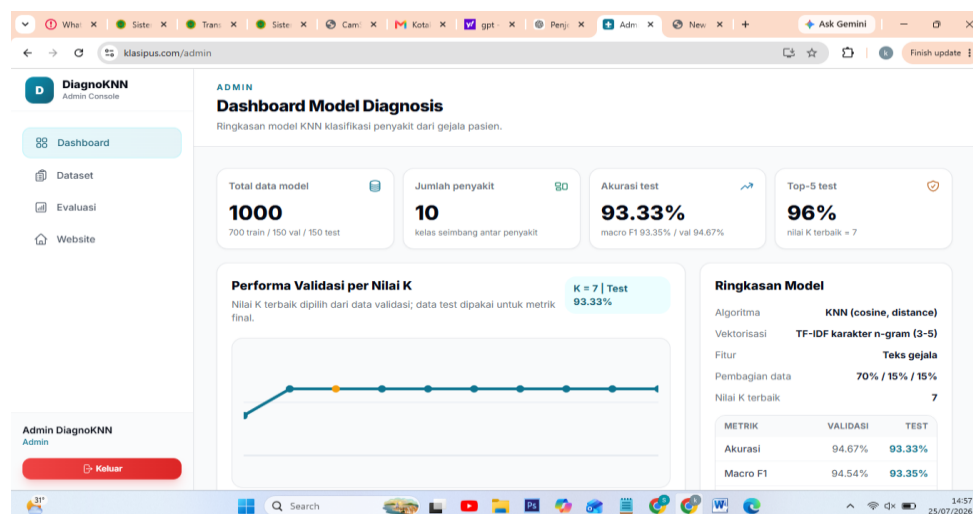


Gambar 7 Login Admin

Gambar tersebut menampilkan halaman Login Admin pada aplikasi DiagnokNN yang berfungsi sebagai gerbang akses bagi administrator untuk masuk ke dashboard pengelolaan sistem. Pada bagian atas terdapat logo dan nama aplikasi DiagnokNN dengan keterangan bahwa dashboard digunakan untuk evaluasi model dan riwayat prediksi. Di bagian tengah terdapat formulir login yang terdiri dari kolom Email dan Password untuk proses autentikasi

pengguna, serta opsi “Ingat sesi login” yang memungkinkan sistem menyimpan sesi pengguna agar tidak perlu login berulang kali. Tersedia tombol “Masuk Dashboard” yang digunakan untuk mengakses halaman admin setelah data login berhasil diverifikasi. Di bagian bawah terdapat tautan “Kembali ke beranda” yang memungkinkan pengguna kembali ke halaman utama website. Secara keseluruhan, halaman ini memiliki desain yang sederhana, rapi, dan mudah digunakan sehingga memudahkan administrator dalam mengakses fitur pengelolaan data, pelatihan model, dan evaluasi sistem DiagnokNN.

c. Dashboard admin



Gambar 8 Dashboard Admin

Gambar tersebut menampilkan Dashboard Model Diagnosis pada halaman admin aplikasi DiagnokNN, yang berfungsi sebagai pusat

pemantauan hasil pelatihan dan evaluasi model klasifikasi penyakit menggunakan algoritma K-Nearest Neighbors (KNN). Pada bagian atas dashboard ditampilkan empat indikator utama yang memberikan ringkasan performa model. Indikator pertama menunjukkan Total Data Model sebanyak 1.000 data, yang dibagi menjadi 700 data training (70%), 150 data validasi (15%), dan 150 data testing (15%). Pembagian ini dilakukan agar model dapat dilatih, divalidasi, dan diuji secara objektif. Indikator kedua menunjukkan bahwa sistem mampu mengklasifikasikan 10 jenis penyakit dengan distribusi data yang telah diseimbangkan. Selanjutnya, indikator Akurasi Test menampilkan nilai 93,33%, dengan Macro F1 sebesar 93,35% dan akurasi validasi sebesar 94,67%, yang menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik pada data uji. Indikator terakhir menampilkan Top-5 Test sebesar 96% dengan nilai K terbaik adalah 7, yang berarti diagnosis yang benar berhasil muncul pada lima rekomendasi teratas sebanyak 96% dari seluruh data pengujian.

Pada bagian tengah dashboard terdapat grafik Performa Validasi per Nilai K yang digunakan untuk membandingkan hasil evaluasi model pada beberapa nilai K. Grafik menunjukkan bahwa performa model meningkat pada nilai K awal dan kemudian cenderung stabil pada nilai-nilai berikutnya. Titik berwarna oranye menandai $K = 7$ sebagai parameter terbaik yang dipilih berdasarkan hasil validasi, sedangkan kotak informasi di sebelah kanan grafik menampilkan bahwa model dengan $K = 7$ menghasilkan akurasi pengujian

sebesar 93,33%. Grafik ini membantu administrator dalam menentukan nilai K yang paling optimal sebelum model digunakan pada proses prediksi.

Di sisi kanan terdapat panel Ringkasan Model yang menjelaskan konfigurasi model yang digunakan. Sistem menerapkan algoritma KNN dengan Cosine Distance sebagai metode pengukuran kemiripan antar data. Proses ekstraksi fitur menggunakan TF-IDF Character N-Gram (3–5) untuk mengubah teks gejala pasien menjadi representasi numerik yang dapat diproses oleh algoritma. Fitur utama yang digunakan dalam klasifikasi adalah teks gejala, sedangkan pembagian data dilakukan dengan rasio 70% training, 15% validasi, dan 15% testing. Panel ini juga menampilkan perbandingan hasil evaluasi antara data validasi dan data testing, yaitu akurasi validasi sebesar 94,67% dan akurasi testing sebesar 93,33%, Macro F1 validasi sebesar 94,54% dan testing sebesar 93,35%, yang menunjukkan bahwa performa model tetap konsisten pada data yang belum pernah dilihat sebelumnya. Secara keseluruhan, dashboard ini memberikan gambaran menyeluruh mengenai kualitas model KNN, mulai dari jumlah data yang digunakan, proses pemilihan nilai K terbaik, hingga hasil evaluasi akhir, sehingga administrator dapat memastikan bahwa model telah memiliki performa yang baik dan siap digunakan sebagai sistem pendukung klasifikasi diagnosis penyakit berdasarkan gejala pasien.

32

BAB V

Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat diambil beberapa kesimpulan sebagai berikut:

34

142

15

5

23

2

7

1. Metode K-Nearest Neighbors (KNN) dapat diterapkan untuk mengklasifikasikan penyakit pasien rawat jalan berdasarkan data gejala yang dimasukkan ke dalam sistem. Proses klasifikasi dilakukan melalui tahapan reprocessing data, pembentukan fitur menggunakan TF-IDF, perhitungan kemiripan menggunakan cosine distance, hingga penentuan kelas berdasarkan tetangga terdekat. Hasil penelitian menunjukkan bahwa metode KNN mampu mengenali pola kemiripan gejala pasien dengan menghasilkan rekomendasi diagnosis penyakit secara otomatis.
2. Hasil pengujian menunjukkan bahwa model terbaik diperoleh pada nilai $K=7$ dengan nilai Accuracy sebesar 94.67%, Macro F1 sebesar 94.54% dan Top-5 Accuracy sebesar 98%. Nilai tersebut menunjukkan bahwa model mampu melakukan klasifikasi penyakit dengan cukup baik. Tingginya nilai Top-5 Accuracy menunjukkan bahwa diagnosis yang hamper selalu muncul dalam lima rekomendasi teratas yang diberikan sistem
3. Tahapan preprocesing seperti pembersihan data, normalisasi istilah gejala, transofmasi data, dan pembentukan fitur lama sakit terbukti membantu

meningkatkan kualitas data yang digunakan dalam proses klasifikasi. Selain itu, penerapan teknik random undersampling mampu mengurangi dominasi kelas mayoritas yang terdapat pada dataset.

6.2 Saran

Berdasarkan hasil penelitian, penulis memberikan saran sebagai berikut:

1. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan lebih beragam agar model dapat mempelajari pola penyakit dengan lebih baik. Semakin banyak data yang digunakan, maka kemampuan model dalam melakukan klasifikasi diharapkan menjadi lebih akurat.
2. Penelitian ini hanya menggunakan fitur gejala pasien lama sakit sebagai fitur utama. Pada penelitian berikutnya dapat ditambahkan atribut lain seperti usia, suhu tubuh, tekanan darah, riwayat penyakit, hasil pemeriksaan laboratorium, dan faktor kesehatan lainnya untuk meningkatkan performa model.

DAFTAR PUSTAKA

- [1] I. F. Rahman, W. A. Shestia, S. D. Rama, S. Azzahra, and A. A. Dermawan, “Klasifikasi Diagnosa Pasien Di Klinik Sri Dengan Metode Decision Tree,” vol. 9, no. 01, 2024.
- [2] A. Yogianto, A. Homaidi, and Z. Fatah, “G-Tech : Jurnal Teknologi Terapan,” vol. 8, no. 3, pp. 1720–1728, 2024.
- [3] D. D. Putri, M. T. Furqon, and R. S. Perdana, “Klasifikasi Penyakit Kulit Pada Manusia Menggunakan Metode Binary Decision Tree Support Vector Machine (BDT SVM) (Studi Kasus : Puskesmas Dinoyo Kota Malang),” vol. 2, no. 5, pp. 1912–1920, 2018.
- [4] N. R. Muntari and K. H. Hanif, “Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning,” vol. 3, no. 1, pp. 1–6, 2022.
- [5] “Siti Kalimah, Klasifikasi Penyakit Diabetes Menggunakan Metode Decision Tree Dan Random Forest” 2022.
- [6] M. Decision, T. Algoritma, M. Decision, and T. Algoritma, “DI RSUD Ir . Soekarno Kabupaten Sukoharjo Klasifikasi Penyakit Terbanyak Di Rsud Ir . Soekarno Kabupaten Sukoharjo Skripsi,” 2024.
- [7] F. M. Hana, “Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma

Decision Tree C4 . 5,” 2020.

- [8] R. Ridho, T. Informatika, F. Teknik, and U. M. Jakarta, “Klasifikasi Diagnosis Penyakit Covid-19 Menggunakan Metode Decision Tree,” vol. 11, no. 3, pp. 69–75, 2021.
- [9] V. No *et al.*, “Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4 . 5 Dengan Rapidminer,” vol. 5, no. 2, pp. 73–83, 2023.
- [10] Y. Widiastiwi and I. Ernawati, “Klasifikasi Penyakit Batu Ginjal Menggunakan Algoritma Decision Tree C4 . 5 Dengan Membandingkan Hasil Uji Akurasi,” vol. 5, no. 2, pp. 128–135.
- [11] N. Putu, B. Ananta, I. Ayu, and G. Suwiprabayanti, “Klasifikasi Jenis Obat Berdasarkan Gejala yang Dimiliki Pasien Menggunakan Metode K-Nearest Neighbors (KNN),” vol. 2, pp. 617–624, 2024.
- [12] C. Journal, M. A. Bianto, M. T. Informatika, and U. A. Yogyakarta, “Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes,” vol. 6, no. 1, 2019.
- [13] F. Wafiyah, N. Hidayat, and R. S. Perdana, “Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam,” vol. 1, no. 10, pp. 1210–1219, 2017.
- [14] I. F. Rahman, W. A. Shestia, S. D. Rama, S. Azzahra, and A. A. Dermawan,

“Klasifikasi Diagnosa Pasien Di Klinik Sri Dengan Metode Decision Tree,”
vol. 9, no. 01, 2024.

- [15] N. R. Muntiari and K. H. Hanif, “Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning,” vol. 3, no. 1, pp. 1–6, 2022.

LAMPIRAN

