

ANALISIS KELULUSAN SISWA MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR) PADA DATASET "STUDENTS PERFORMANCE"

Zaehol Fatah^{1*}, Hali Mukid²

^{1,2}Universitas Ibrahimy Ibrahimy Pondok Pesantren Salafiyah Syafi'iyah Sukorejo Jl. K.H.R. Syamsul Arifin Sukorejo, Sumberejo, Banyuputih, Situbondo, Indonesia; (0338) 453068

Keywords:

K-Nerst Neighbor;
Data Mining;
Analisis Kelulusan.

Correspondent Email:

halimukid79@gmail.com

Abstrak. Tingkat kelulusan siswa menjadi salah satu indikator penting dalam menilai mutu pendidikan. Prediksi yang berkaitan dengan kelulusan dapat mendukung lembaga pendidikan dalam melakukan tindakan pencegahan serta perbaikan dalam proses pengajaran. Penelitian ini memiliki tujuan untuk mengevaluasi efektivitas metode K-Nearest Neighbor (KNN) dalam meramalkan kelulusan siswa dengan menggunakan dataset Kinerja Siswa.. Dataset ini berisi informasi mengenai karakteristik demografis dan akademik, termasuk jenis kelamin, ras, latar belakang pendidikan orang tua, status keikutsertaan dalam program makan siang, program persiapan ujian, dan hasil ujian dalam bidang matematika, membaca, dan menulis. Status kelulusan ditentukan berdasarkan rata-rata nilai ketiga mata pelajaran tersebut. Hasil penelitian menunjukkan bahwa metode KNN mampu memberikan tingkat akurasi yang cukup baik dalam klasifikasi kelulusan siswa. Hasil ini diharapkan bisa menjadi acuan dalam merancang sistem yang membantu pengambilan keputusan dengan memanfaatkan data untuk meningkatkan efisiensi pendidikan.

Abstract. The successful completion of studies by students serves as a key metric when evaluating the standard of educational programs. Anticipating graduation rates can assist schools in implementing proactive strategies and enhancing both teaching and learning experiences. This study aims to determine the effectiveness of the K-Nearest Neighbor (KNN) algorithm in predicting student graduation rates by utilizing the Students Performance dataset. The dataset includes demographic and academic variables, such as gender, ethnicity, parental education, lunch status, exam preparation program, as well as exam scores in mathematics, reading, and writing. Graduation status is determined based on the average score of these three subjects. The findings of the research indicate that the KNN technique yields a reasonably high accuracy rate in categorizing student graduation. These results are anticipated to be a useful guide in creating data-oriented decision support systems aimed at improving the efficiency of educational outcomes.



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

1. PENDAHULUAN

Pendidikan dipandang sebagai salah satu fondasi fundamental dalam mengembangkan tenaga kerja manusia yang berkualitas dan memiliki daya saing. [1]. Kesuksesan dalam sistem pendidikan tidak hanya dilihat dari hasil akademis, tetapi juga dari kemampuan lembaga pendidikan untuk menjamin bahwa setiap siswa memenuhi tolok ukur kompetensi yang telah ditentukan.

Tingkat kelulusan siswa atau rasio keberhasilan adalah salah satu ukuran utama untuk menilai seberapa efektif pendidikan yang disediakan. Tingkat yang tinggi menunjukkan bahwa proses belajar mengajar berjalan dengan baik, kurikulum yang diterapkan sesuai, dan intervensi pendidikan dari lembaga berjalan efektif.[2]. Namun, masih ada sejumlah siswa yang berisiko gagal karena berbagai faktor, seperti kurangnya partisipasi dalam proses pembelajaran, kesulitan memahami materi, ketidakhadiran, di samping faktor sosial dan psikologis yang memengaruhi kinerja akademik[3].

Dalam konteks ini, identifikasi dini siswa yang berpotensi berisiko tidak lulus menjadi sangat penting. Deteksi dini memungkinkan institusi pendidikan untuk mengambil langkah-langkah intervensi preventif, seperti memberikan bimbingan akademik tambahan[4], melaksanakan program remedial, meningkatkan pemantauan melalui konselor akademik, dan memodifikasi strategi pembelajaran[5]. Penerapan sistem prediksi keberhasilan sebagai bagian dari Sistem Peringatan Dini (EWS) merupakan strategi krusial bagi institusi.

Kemajuan signifikan dalam bidang Penambangan Data Pendidikan (Educational Data Mining - EDM) dan Pembelajaran Mesin (Machine Learning) telah menghasilkan kemampuan prediksi yang kuat dan akurat. Model seperti Hutan Acak (Random Forest) dan lainnya telah terbukti efektif.

Model K-Nearest Neighbors (KNN) adalah salah satu teknik yang paling sering

digunakan dalam penelitian prediksi di area akademis. karena pendekatannya yang intuitif, tidak bergantung pada asumsi distribusi data, dan kemampuannya untuk beradaptasi dengan variasi pola data setelah langkah-langkah pra-pemrosesan seperti normalisasi, reduksi dimensi, dan penyeimbangan kelas[6]. Menunjukkan bahwa model KNN mampu memprediksi penerimaan siswa dengan akurasi tinggi setelah pemilihan fitur dan pra-pemrosesan yang memadai. Pada saat yang sama[7].

KNN memiliki kinerja yang baik dalam mengklasifikasikan status keberhasilan mahasiswa di Politeknik Negeri Tanah Laut[8]. Hasil-hasil ini memperkuat bukti bahwa KNN adalah algoritma yang andal dan cocok untuk digunakan dalam konteks prediksi akademik.

Mengingat urgensi identifikasi dini risiko ketidakhadiran dan efektivitas metode KNN berdasarkan bukti empiris, penelitian ini berfokus pada penerapan algoritma KNN untuk memprediksi status keberhasilan siswa menggunakan data akademik yang tersedia[9]. Hasil yang didapat diharapkan bisa memberikan sumbangsih pada pengembangan Sistem Pendukung Keputusan (DSS) yang menggunakan Penambangan Data Pendidikan (EDM), yang dapat membantu lembaga pendidikan dalam merencanakan intervensi.

Oleh karena itu, penerapan metode KNN dalam konteks Penambangan Data Pendidikan merupakan salah satu strategi yang menjanjikan untuk mendukung sistem peringatan dini akademik serta pengambilan keputusan yang berbasis data di bidang pendidikan.

2. TINJAUAN PUSTAKA

Data Mining

Pengolahan informasi merupakan tahap dalam menganalisis sekumpulan data yang luas dan rumit, dengan niatan untuk mengidentifikasi pola atau fakta signifikan yang sebelumnya tidak terdeteksi. Istilah pengolahan data sering berhubungan dengan identifikasi pola dan penemuan pengetahuan.[10]

Pengenalan pola atau pattern recognition diterapkan karena hasil utama dari proses ini adalah pola-pola tersembunyi yang dapat dikenali dari data dalam skala besar. Di sisi lain, istilah penemuan pengetahuan atau knowledge discovery juga berkaitan, karena inti dari penggalian data adalah menemukan dan menarik pengetahuan baru dari kumpulan data yang besar. Dalam penelitian ini, istilah *data mining* dipilih karena lebih umum digunakan dalam konteks penelitian ilmiah dibandingkan istilah lainnya[11].

Algoritma K-Nearest Neighbor (KNN) beroperasi dengan cara mengukur jarak antara dua titik data, yaitu data yang sedang diuji dan data yang telah dilatih. Setelah jarak antar titik dihitung, algoritma akan menentukan kelas suatu data baru berdasarkan mayoritas dari k data latih yang memiliki jarak terdekat dengannya[12].

Dalam penelitian ini, perhitungan jarak antar data menggunakan rumus **Euclidean Distance**, yang secara matematis dapat dinyatakan pada Persamaan (3):

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

Keterangan:

- d = jarak yang memisahkan data pelatihan dan data pengujian
- i = indeks dari atribut yang ke- i
- n = total atribut (dimensi dari data)
- x_i = nilai untuk atribut yang ke- i pada data pelatihan
- y_i = nilai atribut yang ke- i dalam data pengujian

Pada algoritma ini dikembangkan berdasarkan asumsi bahwa data yang memiliki kedekatan jarak dalam ruang fitur cenderung berada pada kelas yang sama.

3. METODE PENELITIAN

3.1 Sumber Data

Adapun data ini diperoleh dari web keggel.com yang berupa Dataset

“**StudentsPerformance**” berisi data tentang performa akademik para siswa beserta faktor-faktor yang mungkin memengaruhi hasil belajar. Teks ini berisi data seperti gender, suku atau kelompok etnis, pendidikan orang tua, serta jenis makanan siang yang diterima oleh siswa, serta apakah siswa mengikuti kursus persiapan tes atau tidak. Selain itu, terdapat tiga nilai utama yang menggambarkan kemampuan akademik siswa, yaitu nilai matematika, membaca, dan menulis. Dari ketiga nilai itu, dihitunglah rata-rata, yang selanjutnya dipakai untuk menentukan apakah seorang siswa berhasil lulus atau tidak.

Secara keseluruhan, dataset ini menggambarkan hubungan antara latar belakang keluarga, kondisi sosial, dan persiapan belajar siswa dengan hasil akademik. Misalnya, jenis makan siang dapat menjadi indikator kondisi ekonomi, pendidikan orang tua dapat menunjukkan dukungan belajar di rumah, dan kursus persiapan dapat memengaruhi nilai siswa. Data semacam ini umumnya dimanfaatkan untuk menyelidiki elemen-elemen apa saja yang paling berdampak pada kinerja siswa, menghasilkan gambaran pola nilai, serta menciptakan model untuk meramalkan kelulusan berdasarkan variabel yang ada.[13].

student_id	gender	race/ethnicity	lunch	test_preparation	math_score	reading_score	writing_score
1	female	group A	standard	none	54	58	54
2	female	group A	standard	none	47	54	47
3	female	group A	standard	none	49	49	51
4	female	group A	standard	none	51	51	51
5	female	group A	standard	none	53	53	53
6	female	group A	standard	none	55	55	55
7	female	group A	standard	none	57	57	57
8	female	group A	standard	none	59	59	59
9	female	group A	standard	none	61	61	61
10	female	group A	standard	none	63	63	63
11	female	group A	standard	none	65	65	65
12	female	group A	standard	none	67	67	67
13	female	group A	standard	none	69	69	69
14	female	group A	standard	none	71	71	71
15	female	group A	standard	none	73	73	73
16	female	group A	standard	none	75	75	75
17	female	group A	standard	none	77	77	77
18	female	group A	standard	none	79	79	79
19	female	group A	standard	none	81	81	81
20	female	group A	standard	none	83	83	83
21	female	group A	standard	none	85	85	85
22	female	group A	standard	none	87	87	87
23	female	group A	standard	none	89	89	89
24	female	group A	standard	none	91	91	91
25	female	group A	standard	none	93	93	93
26	female	group A	standard	none	95	95	95
27	female	group A	standard	none	97	97	97
28	female	group A	standard	none	99	99	99
29	female	group A	standard	none	100	100	100
30	female	group A	standard	none	100	100	100
31	female	group A	standard	none	100	100	100
32	female	group A	standard	none	100	100	100
33	female	group A	standard	none	100	100	100
34	female	group A	standard	none	100	100	100
35	female	group A	standard	none	100	100	100
36	female	group A	standard	none	100	100	100
37	female	group A	standard	none	100	100	100
38	female	group A	standard	none	100	100	100
39	female	group A	standard	none	100	100	100
40	female	group A	standard	none	100	100	100
41	female	group A	standard	none	100	100	100
42	female	group A	standard	none	100	100	100
43	female	group A	standard	none	100	100	100
44	female	group A	standard	none	100	100	100
45	female	group A	standard	none	100	100	100
46	female	group A	standard	none	100	100	100
47	female	group A	standard	none	100	100	100
48	female	group A	standard	none	100	100	100
49	female	group A	standard	none	100	100	100
50	female	group A	standard	none	100	100	100
51	female	group A	standard	none	100	100	100
52	female	group A	standard	none	100	100	100
53	female	group A	standard	none	100	100	100
54	female	group A	standard	none	100	100	100
55	female	group A	standard	none	100	100	100
56	female	group A	standard	none	100	100	100
57	female	group A	standard	none	100	100	100
58	female	group A	standard	none	100	100	100
59	female	group A	standard	none	100	100	100
60	female	group A	standard	none	100	100	100
61	female	group A	standard	none	100	100	100
62	female	group A	standard	none	100	100	100
63	female	group A	standard	none	100	100	100
64	female	group A	standard	none	100	100	100
65	female	group A	standard	none	100	100	100
66	female	group A	standard	none	100	100	100
67	female	group A	standard	none	100	100	100
68	female	group A	standard	none	100	100	100
69	female	group A	standard	none	100	100	100
70	female	group A	standard	none	100	100	100
71	female	group A	standard	none	100	100	100
72	female	group A	standard	none	100	100	100
73	female	group A	standard	none	100	100	100
74	female	group A	standard	none	100	100	100
75	female	group A	standard	none	100	100	100
76	female	group A	standard	none	100	100	100
77	female	group A	standard	none	100	100	100
78	female	group A	standard	none	100	100	100
79	female	group A	standard	none	100	100	100
80	female	group A	standard	none	100	100	100
81	female	group A	standard	none	100	100	100
82	female	group A	standard	none	100	100	100
83	female	group A	standard	none	100	100	100
84	female	group A	standard	none	100	100	100
85	female	group A	standard	none	100	100	100
86	female	group A	standard	none	100	100	100
87	female	group A	standard	none	100	100	100
88	female	group A	standard	none	100	100	100
89	female	group A	standard	none	100	100	100
90	female	group A	standard	none	100	100	100
91	female	group A	standard	none	100	100	100
92	female	group A	standard	none	100	100	100
93	female	group A	standard	none	100	100	100
94	female	group A	standard	none	100	100	100
95	female	group A	standard	none	100	100	100
96	female	group A	standard	none	100	100	100
97	female	group A	standard	none	100	100	100
98	female	group A	standard	none	100	100	100
99	female	group A	standard	none	100	100	100
100	female	group A	standard	none	100	100	100

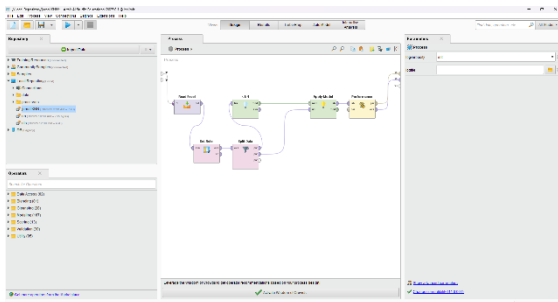
Gambar 3.1. Dataset Penelitian

3.2 Preprocessing Data

Pada fase persiapan data, dilakukan beberapa tindakan untuk memastikan bahwa data sudah siap untuk digunakan oleh model machine learning. Tindakan awal adalah mengonversi semua data kategorikal menjadi bentuk angka agar bisa diproses oleh algoritma. Hal ini dilakukan dengan menggunakan teknik

label encoding, yang memberikan representasi angka pada kategori tertentu, misalnya mengubah label “male” dan “female” menjadi angka, begitu juga dengan kategori pada variabel makan siang serta kelompok etnis.

Setelah itu, dilakukan perhitungan **rata-rata nilai** dari tiga mata pelajaran, yaitu matematika, membaca, dan menulis. Rata-rata ini menjadi indikator utama yang digunakan untuk menentukan status kelulusan siswa. Berdasarkan nilai rata-rata tersebut, dibuatlah **label klasifikasi**, di mana siswa dengan nilai rata-rata **70 atau lebih** dikategorikan sebagai **“Lulus”**, sedangkan siswa dengan nilai rata-rata **di bawah 70** dimasukkan ke kategori **“Tidak Lulus”**. Klasifikasi ini penting karena menjadi target output yang akan dipelajari oleh model.



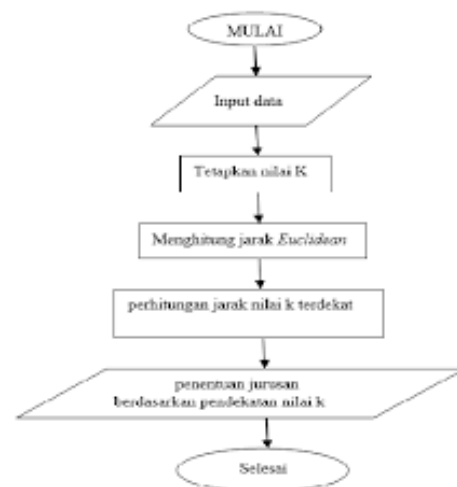
Gambar 3.2 Preprocessing Data

Terakhir, agar dapat menilai model secara akurat dan memahami seberapa efektif kemampuannya dalam beradaptasi, informasi dibagi menjadi dua kelompok, yaitu informasi yang dimanfaatkan untuk pelatihan dan informasi untuk evaluasi. Pembagian ini dilakukan secara acak dengan proporsi 80% untuk pelatihan dan 20% untuk evaluasi. Data yang dipakai pada tahap pelatihan berfungsi untuk merancang dan mengatur parameter model, sementara itu, informasi yang diterapkan untuk evaluasi bertujuan untuk menilai seberapa efektif model dalam mengidentifikasi pola dari data yang segar dan tidak familiar. Pemisahan ini adalah prosedur umum dalam pembelajaran mesin agar menghindari overfitting dan menjamin model dapat berfungsi dengan baik di luar data pelatihan.

3.3 Algoritma Klasifikasi

Pada bagian ini dijelaskan algoritma klasifikasi yang digunakan dalam penelitian ini, termasuk algoritma K-Nearest Neighbor (KNN). Algoritma ini dipandang sebagai salah satu cara yang mudah dan efisien untuk mengelompokkan data berdasar pada sejauh mana fitur-fitur tersebut memiliki kesamaan. Algoritma ini didasarkan pada prinsip bahwa titik data baru akan diklasifikasikan ke dalam kategori yang sama dengan mayoritas titik terdekat dalam ruang fitur. Untuk menentukan tingkat kedekatan antar titik, penelitian ini menggunakan ukuran jarak Euclidean, yang mengukur jarak antara setiap titik data dengan titik-titik data pada kumpulan pelatihan[14].

Proses pengelompokan dalam teknik KNN dimulai dengan menentukan nilai k , yaitu total tetangga terdekat yang akan diperhatikan untuk menetapkan klasifikasi. Pemilihan angka k yang sesuai sangat krusial karena angka ini berpengaruh besar terhadap ketepatan hasil; angka k yang terlalu rendah dapat membuat model terlalu peka terhadap gangguan, sedangkan angka yang terlalu tinggi dapat menghasilkan keputusan yang terlalu umum.[15]. Setelah nilai k ditetapkan, sistem menganalisa perbedaan antara data uji dan seluruh data pelatihan dengan menggunakan rumus jarak Euclidean. Data pelatihan kemudian diurutkan secara naik berdasarkan jarak tersebut, dan k titik terdekat dipilih untuk menentukan kategori yang sesuai bagi data uji.



Gambar 3.3 Alur Klasifikasi

Melalui pendekatan ini, algoritma KNN mampu menghasilkan hasil klasifikasi yang akurat, terutama ketika data memiliki pola distribusi yang jelas. Berkat sifatnya yang berbasis contoh dan tidak memerlukan proses pelatihan yang kompleks, algoritma ini dipilih dalam penelitian untuk mempermudah proses penentuan kategori data yang dianalisis.

3.4 Tool dan Platform

Dalam studi ini, alat yang dipakai untuk melaksanakan pengolahan data mining adalah RapidMiner. RapidMiner merupakan salah satu alat terkemuka untuk analisis data dan machine learning, yang dirancang untuk membantu pengguna membangun model analitik tanpa perlu melakukan pemrograman. Perangkat lunak ini memiliki antarmuka grafis berbasis sistem drag-and-drop, yang memungkinkan pengguna membuat alur kerja analitik secara intuitif hanya dengan menghubungkan operator-operator yang tersedia di dalam lingkungan kerja.

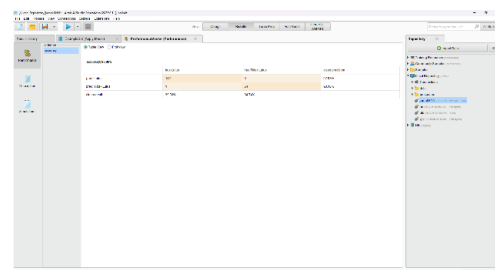
RapidMiner menawarkan lingkungan kerja yang lengkap dan mendukung seluruh tahapan proses data mining, mulai dari impor dan pembersihan data, diikuti oleh transformasi serta praproses data, hingga pemilihan algoritma dan evaluasi model klasifikasi atau prediksi. Alur kerja yang terstruktur ini membuat proses analisis menjadi lebih teratur dan mudah diikuti. Selain itu, RapidMiner menyediakan berbagai pilihan algoritma pembelajaran mesin seperti K-Nearest Neighbor (KNN), Pohon Keputusan, Naive Bayes, Support Vector Machine (SVM), serta teknik klastering dan regresi. Variasi ini memberi kesempatan kepada peneliti untuk dengan mudah membandingkan performa berbagai algoritma hanya dengan mengganti operator dalam proses kerja.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pengujian K-Nearest Neighbor (KNN)

Evaluasi dilakukan dengan menggunakan *Performance Vector*, yang menampilkan model akurasi, matriks konfusi, serta nilai presisi dan recall untuk setiap kelas. Berdasarkan hasil pengolahan, model KNN

menghasilkan akurasi sebesar 98%. Akurasi ini dihitung dengan mempertimbangkan rasio antara jumlah prediksi yang benar dan total data uji secara keseluruhan. Nilai akurasi sebesar 98% menunjukkan bahwa model berhasil dalam mengidentifikasi pola data dengan sangat baik dan hanya melakukan kesalahan pada sebagian kecil data. Selain nilai akurasi, hasil evaluasi model juga disajikan melalui matriks konfusi, yang memperlihatkan perbandingan antara kelas yang sebenarnya dan kelas hasil prediksi. Berdasarkan matriks tersebut, terdapat 142 data *Lulus* yang berhasil diprediksi dengan benar, dan 54 data *Tidak Lulus* yang juga diprediksi dengan benar. Sementara itu, kesalahan prediksi hanya terjadi pada 4 data, yaitu 3 data *Tidak Lulus* diprediksi sebagai *Lulus*, dan 1 data *Lulus* diprediksi menjadi *Tidak Lulus*. Hasil ini menunjukkan bahwa tingkat kesalahan model sangat rendah. Selanjutnya, RapidMiner juga menampilkan nilai presisi dan recall untuk masing-masing kelas. Presisi untuk kelas *Lulus* adalah 97.93%, sedangkan presisi untuk kelas *Tidak Lulus* mencapai 98.18%. Angka ini menunjukkan bahwa sebagian besar data yang diprediksi oleh model benar sesuai kelasnya. Sementara itu, tingkat recall untuk kategori *Lulus* mencapai angka 99.30%, sedangkan untuk kategori *Tidak Lulus* adalah 94.74%, yang menunjukkan kemampuan model dalam mengenali setiap kategori dengan sangat efektif. Data latihan digunakan untuk merancang serta menyesuaikan parameter dari model, sedangkan data uji dimanfaatkan untuk mengevaluasi seberapa baik model dapat mengidentifikasi pola dalam data yang baru dan belum pernah ditemui sebelumnya. Pemisahan ini adalah prosedur umum dalam pembelajaran mesin agar menghindari overfitting dan menjamin model dapat berfungsi dengan baik di luar data pelatihan.



Gambar 4. 1 Hasil Pengujian K-Nearest Neighbor (KNN)

4.2 Pembahasan

Hasil uji model K-Nearest Neighbors (KNN) dalam studi ini menunjukkan bahwa algoritma tersebut mampu memberikan kinerja klasifikasi yang sangat kuat untuk data kelulusan mahasiswa. Akurasinya mencapai 98%, yang menunjukkan bahwa sebagian besar data pengujian berhasil diprediksi dengan benar oleh model. Akurasi yang tinggi ini mencerminkan kemampuan algoritma dalam mengenali pola secara efektif antara variabel masukan dan label kelas dalam kumpulan data.

The screenshot shows the WEKA Explorer interface with the 'Confusion Matrix' tab selected. The matrix displays counts for 'Actual' vs 'Predicted' classes, with a high accuracy of 98%.

	Actual \ Predicted	Benar	Salah	Total
Benar	98	2	100	
Salah	2	98	100	
Total	100	100	200	

Gambar 4.2 Hasil Pengujian K-Nearest Neighbor

Dalam evaluasi menggunakan confusion matrix menunjukkan bahwa model hanya melakukan empat kesalahan, terdiri dari tiga kesalahan prediksi pada kelas “tidak lulus” dan satu kesalahan pada kelas “lulus”. Kesalahan ini mungkin disebabkan oleh perbedaan distribusi data antara kedua kelas tersebut. Namun, jumlah kesalahan yang sangat sedikit menunjukkan bahwa model tidak mengalami masalah serius seperti overfitting atau underfitting. Dengan demikian, model KNN terbukti efektif dalam memanfaatkan prinsip kedekatan antar sampel di ruang fitur, sehingga memungkinkan model memberikan prediksi yang akurat.

Selain itu, nilai precision dan recall untuk setiap kelas menunjukkan kinerja yang stabil dan konsisten. Precision untuk kelas lulus dan tidak lulus masing-masing melebihi 97%, yang berarti sebagian besar prediksi model sesuai dengan kenyataan. Nilai recall yang tinggi, terutama pada kelas lulus (99,30%), menunjukkan kemampuan model dalam mengenali sebagian besar data yang termasuk

dalam kelas tersebut. Meskipun recall pada kelas tidak lulus sedikit lebih rendah (94,74%), hasil ini tetap sangat baik dan mencerminkan efektivitas model dalam mendeteksi kedua kelas.

Secara keseluruhan, kombinasi antara akurasi yang tinggi, precision dan recall yang konsisten, serta jumlah kesalahan yang sangat sedikit pada confusion matrix menunjukkan bahwa algoritma KNN sangat cocok untuk menyelesaikan permasalahan klasifikasi dalam studi ini. Model ini mampu beradaptasi dengan baik terhadap pola data melalui prinsip kedekatan antar sampel, yang menjadi dasar dari metode ini. Oleh karena itu, model ini merupakan pilihan yang sangat berguna dan dapat diandalkan dalam mendukung pengambilan keputusan atau sistem pendukung keputusan terkait tugas prediksi.

5. KESIMPULAN

Dalam penelitian ini dapat disimpulkan bahwa :

- Model K-Nearest Neighbors (KNN) berhasil mencapai tingkat ketepatan hingga 98%. Akurasi ini mengukur proporsi total prediksi yang benar dari seluruh data uji. Tingkat akurasi yang mendekati sempurna (98%) menunjukkan bahwa algoritma KNN sangat efektif dan mampu mengklasifikasi sebagian besar data, menjadikannya model yang sangat dapat diandalkan untuk memprediksi hasil *manh* (pemberian/hibah).
- Berdasarkan Matriks Kebingungan (Confusion Matrix), model hanya menghasilkan empat (4) kesalahan prediksi. Kesalahan ini terdiri dari 3 kasus "Tidak Berhasil" yang salah diprediksi sebagai "Berhasil" (*False Positive*) dan 1 kasus "Berhasil" yang salah diprediksi sebagai "Tidak Berhasil" (*False Negative*). Jumlah kesalahan yang sangat kecil ini adalah indikator kuat stabilitas model. Sedikitnya kesalahan menunjukkan bahwa model mampu mempelajari pola data dengan baik tanpa menderita masalah *overfitting* (terlalu terpaku

pada data pelatihan) atau *underfitting* (gagal menangkap pola data).

DAFTAR PUSTAKA

- [1] P. Berkualitas, “Mengoptimalkan potensi sumber daya manusia untuk mewujudkan pendidikan berkualitas,” vol. 6, no. 2, pp. 641–650, 2025.
- [2] F. Kobandaha and A. N. Annas, “Pembelajaran Inovatif: Studi Literatur Tentang Faktor yang Mempengaruhi Keberhasilan Peserta Didik,” pp. 137–145.
- [3] J. Lemantara and T. Lusiani, “Analisis Prediksi Penyakit Diabetes Pada Wanita Menggunakan Metode Naïve Bayes Dan K-Nearest Neighbor,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4911.
- [4] F. Syariah and D. I. Iain, “HUKUM KELUARGA ISLAM PROGRAM STUDI HUKUM KELUARGA ISLAM (AL-FAKULTAS SYARIAH UNIVERSITAS NEGERI (UIN) PALOPO PROGRAM STUDI HUKUM KELUARGA ISLAM (AL-FAKULTAS SYARIAH UNVERSITAS ISLAM NEGERI (UIN) PALOPO,” 2025.
- [5] M. M. Baharuddin, H. Azis, and T. Hasanuddin, “Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca,” *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
- [6] Z. Amri, K. Kusriani, and K. Kusnawi, “Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM,” *Edumatic J. Pendidik. Inform.*, vol. 7, no. 2, pp. 187–196, 2023, doi: 10.29408/edumatic.v7i2.18620.
- [7] D. D. Putri, G. F. Nama, and W. E. Sulistiono, “ANALISIS SENTIMEN KINERJA DEWAN PERWAKILAN RAKYAT (DPR) PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER,” vol. 10, no. 1, pp. 34–40, 2022.
- [8] H. M. Nur, V. Maarif, F. Fandi, and D. Imaniawan, “Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Ketidakkuluran Mahasiswa Berdasar Presensi dan IPK,” vol. 4, no. 1, pp. 1–6, 2025.
- [9] D. Pendidikan, “Rahady Puji Alfiansyah Fakultas Teknik Industri dan Informatika, Universitas Muhammadiyah Prof.Dr.Hamka, Indonesia,” vol. 3, no. 6, pp. 469–473, 2023.
- [10] X. Shu and Y. Ye, “Knowledge Discovery : Methods from data mining and machine learning ☆,” *Soc. Sci. Res.*, vol. 110, no. October 2022, p. 102817, 2023, doi: 10.1016/j.ssresearch.2022.102817.
- [11] S. Widaningsih, A. Suheri, and T. T. Fauziyana, “Penerapan Data Mining Menggunakan Algoritma K- Nearest Neighbor untuk Penentuan Penerimaan Proposal Hibah Application of Data Mining Using The K-Nearest Neighbor Algorithm for Determining Acceptance of Grand Proposals,” vol. 8, no. 1, pp. 9–18, 2024, doi: 10.35194/jmtsi.v8i1.2696.
- [12] M. A. Aprihartha *et al.*, “Perbandingan Metrik Euclidean dan Metrik Manhattan untuk K-Nearest Neighbors dalam Klasifikasi Kismis pengolahannya seperti pengeringan , grading , dll . Pada beberapa tempat , penyortiran kismis masih alternatif dalam mengidentifikasi kualitas kismis . Beberapa penelitian terdahulu telah melakukan studi fleksibel . Metode ini bekerja dengan memperhatikan kesamaan karakteristik pada data terdekatnya . pada metrik Manhattan , yaitu 84 % . Penelitian oleh Salsabila , et . al [9] yang membandingkan metrik diperoleh akurasi metrik Manhattan lebih tinggi daripada metrik Euclidean . Meskipun dari dua sampel cukup bukti bahwa hasil serupa dapat terjadi pada studi kasus yang berbeda . Dalam literatur yang ada , dan adanya pustaka (library) pada software R . metode memiliki perbedaan dalam mengukur jarak antar dua atau lebih objek . Setelah dilakukan analisis penilaian akurasi , sensitivitas , spesifisitas , dan presisi . Hasil penelitian ini diharapkan dapat menjadi Dataset terdiri dari 8 variabel dengan jumlah sampel sebanyak 900 amatan .,” vol. 4, no. 1, pp. 21–30, 2024.
- [13] S. Kasus and S. Karawang, “Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama,” vol. 9, no. 1, pp. 14–19, 2023.
- [14] J. Penerapan, T. Informasi, N. T. Ujianto, H. Fadilah, A. P. Fanti, and A. D. Saputra, “IT-EXPLORE Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis,” vol. 02, pp. 33–43, 2025.
- [15] “No Title,” *Aswir, E. S., Wasilah, W. (2024). Implementasi YOLOv3 Menggunakan Fitur Ekstraktor ResNeXt Untuk Deteksi Filariasis. J. Pepadun, 5(2), 131-140.*