

Penerapan Algoritma K-Nearest Neighbor (K-NN) untuk Memprediksi Kelulusan Mahasiswa Menggunakan RapidMiner

Zaehol Fatah¹, Herlinatus Safira Muasolli²

^{1,2} Sistem Informasi, Universitas Ibrahimy Sukorejo

Email : ¹zaeholfatah@mail.com, ^{2*}herlinatussafiramuasolli@gmail.com

Abstrak

Prediksi kelulusan mahasiswa dapat membantu perguruan tinggi dalam meningkatkan mutu dan efisiensi penyelenggaraan pendidikan. Penelitian ini bertujuan untuk membangun model klasifikasi kelulusan mahasiswa dengan bantuan metode K-Nearest Neighbor yang diterapkan melalui perangkat lunak RapidMiner. Data yang digunakan mencakup berbagai atribut akademik dan administratif mahasiswa, dan telah melalui tahap prapemrosesan seperti normalisasi dan penetapan label. Hasil menunjukkan bahwa algoritma K-NN mampu mencapai akurasi sebesar 95,88%, terutama fokus pada klasifikasi mahasiswa dengan kelulusan sesuai waktu yang ditentukan, hal ini membuktikan bahwa algoritma K-NN mampu digunakan sebagai perangkat yang berperan signifikan dalam proses evaluasi dan pengambilan keputusan akademik

Kata Kunci: Klasifikasi, K-NN, RapidMiner, Kelulusan Mahasiswa, Data Mining

Abstrak

Student graduation prediction can help universities improve the quality and efficiency of education delivery. This study aims to build a classification model using the K-Nearest Neighbor (K-NN) method implemented through RapidMiner software. The dataset consists of various academic and administrative attributes of students and has been processed through several stages, including normalization and labeling. The results show that the K-NN algorithm achieves an accuracy of 95.88%, particularly in classifying students who graduate within the specified time, indicating that this algorithm can serve as a reliable tool to support evaluation processes and academic decision-making.

Keywords: Classification, K-NN, RapidMiner, Student Graduation, Data Mining

PENDAHULUAN

Sejalan dengan pesatnya perkembangan era modern, teknologi informasi mengalami kemajuan yang signifikan. Berbagai sumber data dari sektor ekonomi, industri, ilmu pengetahuan, dan teknologi kini dapat dihasilkan melalui pemanfaatan teknologi informasi yang semakin canggih. Dalam bidang pendidikan, teknologi informasi berperan penting dalam menghasilkan data yang melimpah, baik mengenai peserta didik maupun individu lainnya. Khususnya di perguruan tinggi, penerapan teknologi informasi menjadi aspek krusial untuk membantu pengelolaan data akademik. Hal

ini terjadi karena jumlah mahasiswa terus meningkat setiap tahun, sehingga institusi pendidikan tinggi harus mampu mengelola dan memanfaatkan data akademik yang semakin besar, seperti data kelulusan, profil mahasiswa, serta capaian akademik selama proses pendidikan.

Prediksi akurasi kelulusan mahasiswa dibuat untuk membantu program studi dalam membimbing mahasiswa agar lulus sesuai jadwal. Dengan memahami prediksi status kelulusan selama perkuliahan, program studi bersama dosen pembimbing akademik dapat memberikan perhatian khusus kepada mahasiswa yang diperkirakan tidak

menyelesaikan studi tepat waktu. Dengan demikian, mahasiswa tersebut dapat meningkatkan performa akademiknya tiap semester agar lulus sesuai target waktu.

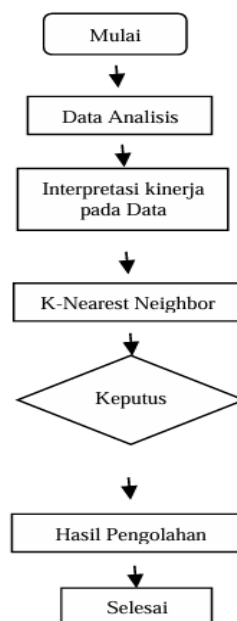
Algoritma yang digunakan untuk menjalankan fungsi klasifikasi adalah K-NN. Algoritma *k-Nearest Neighbors* merupakan metode pembelajaran terawasi, di mana instance baru diklasifikasikan sesuai kategori yang paling dominan dari k tetangga terdekat.

Menurut Pandi, algoritma K-Nearest Neighbors (KNN) mengklasifikasikan data baru berdasarkan jarak terdekat dengan data yang sudah ada. Jumlah tetangga terdekat ditentukan oleh nilai k . Untuk mendapatkan nilai k yang optimal, digunakan metode *k-Fold Cross Validation* yang menguji akurasi sistem melalui pengacakan dan pengujian ulang data. Nilai $k = 1$ bukan satu-satunya pilihan terbaik, karena hasil prediksi sangat dipengaruhi oleh satu data terdekat saja. Oleh karena itu, pemilihan k terbaik didasarkan pada akurasi hasil validasi.

Berdasarkan latar belakang tersebut, penulis berinisiatif melakukan pengujian algoritma K-NN. Penerapan metode klasifikasi ini memungkinkan algoritma K-NN digunakan dalam menentukan kemungkinan kelulusan mahasiswa sehingga diharapkan dapat membantu mereka lulus tepat waktu. Namun, penelitian sebelumnya sebagian besar hanya menekankan aspek teknis penerapan algoritma atau perbandingan metode, tanpa membahas akurasi prediksi dan kontribusinya terhadap evaluasi akademik. Penelitian ini hadir untuk mengisi kekosongan tersebut dengan menghadirkan model klasifikasi menggunakan algoritma K-NN melalui RapidMiner, yang menitikberatkan pada capaian akurasi serta manfaat praktisnya dalam mendukung pengambilan keputusan di perguruan tinggi.

METODE

Gambar tersebut memperlihatkan urutan tahapan penelitian yang dilakukan dengan memanfaatkan algoritma K-NN



Gambar 1. Tahapan Metode Penelitian

1. Mulai

Tahapan pertama adalah inisiasi penelitian.

Pada tahap ini peneliti menentukan tujuan penelitian, objek yang akan diteliti (misalnya data mahasiswa), serta masalah yang ingin diselesaikan menggunakan metode K-NN. Peneliti juga menyiapkan kebutuhan awal seperti perangkat lunak (misalnya RapidMiner) dan sumber data.

2. Data Analisis

Tahap ini mencakup pengumpulan dan analisis awal terhadap data yang akan digunakan.

Langkah-langkahnya meliputi:

- Mengidentifikasi atribut (variabel) yang relevan, misalnya IPK, kehadiran, lama studi, status skripsi, dsb.
- Menentukan atribut input (independent variable) dan output (label) yang akan diprediksi.
- Melakukan data cleaning untuk menghapus data kosong atau tidak valid agar hasil analisis lebih akurat.

3. Interpretasi Kinerja pada Data

Tahap ini bertujuan untuk memahami karakteristik data dan mempersiapkan data agar sesuai dengan algoritma K-NN.

Langkah-langkahnya mencakup:

- Normalisasi data agar setiap atribut memiliki skala yang seragam.

- Pembagian dataset menjadi data latih (training set) dan data uji (testing set).
- Menganalisis distribusi nilai untuk melihat pola yang mungkin memengaruhi hasil klasifikasi.

4. K-Nearest Neighbor

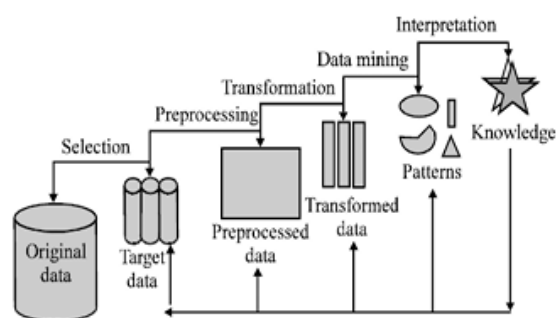
Ini merupakan inti dari penelitian, yaitu penerapan algoritma K-NN.

Tahap ini meliputi:

- Menentukan nilai K (jumlah tetangga terdekat) yang paling optimal.
- Menghitung jarak antar data (biasanya menggunakan Euclidean Distance) antara data uji dan data latih.
- Menentukan kelas data uji berdasarkan mayoritas kelas dari K tetangga terdekat.

Contohnya, jika dari 5 tetangga terdekat ($K=5$), 3 termasuk kategori “Lulus Tepat Waktu”, maka data uji tersebut diklasifikasikan ke dalam kategori yang sama.

Salah satu tahapan penting dalam proses data mining adalah *preprocessing* atau prapemrosesan data. Data yang akan digunakan untuk analisis sering kali tidak berada dalam kondisi ideal. Masalah umum yang ditemukan meliputi data yang tidak lengkap, keberadaan *outlier*, atau format yang tidak konsisten, yang semuanya dapat memengaruhi hasil dari proses analisis data.



Gambar 2. Posisi data preprocessing pada data mining

Gambar tersebut menunjukkan **alur proses dalam Knowledge Discovery in Databases (KDD)**, yaitu tahapan-tahapan untuk mendapatkan pengetahuan dari data mentah hingga menjadi informasi yang bermakna.

Tahapan Preprocessing dalam Data Mining

1. Original Data (Data Asli)

Data mentah yang dikumpulkan dari berbagai sumber, seperti database akademik, formulir mahasiswa, atau sistem informasi kampus.

Data ini masih belum siap digunakan karena bisa saja mengandung noise, duplikasi, atau nilai yang hilang.

2. Selection (Seleksi Data)

Pada tahap ini, hanya data yang relevan dengan tujuan penelitian yang dipilih.

Misalnya, atribut yang berkaitan langsung dengan kelulusan mahasiswa seperti IPK, jumlah SKS, lama studi, dan status kelulusan.

3. Preprocessing (Prapemrosesan Data)

Tahap penting untuk membersihkan dan menyiapkan data agar siap digunakan dalam proses analisis.

Langkah-langkah di dalamnya meliputi:

- Menghapus data kosong atau tidak valid
- Mengatasi outlier
- Menormalkan skala data numerik
- Mengubah format data agar seragam

4. Transformation (Transformasi Data)

Data yang telah dibersihkan kemudian diubah menjadi bentuk yang sesuai untuk algoritma data mining.

Contohnya:

- Mengonversi data kategorikal menjadi numerik
 - Melakukan normalisasi nilai agar tiap atribut memiliki bobot seimbang
 - Membuat atribut baru (feature engineering) bila diperlukan
- #### 5. Data Mining (Proses Penambangan Data)

Setelah data siap, tahap ini menggunakan algoritma tertentu — dalam konteks penelitianmu, algoritma K-Nearest Neighbor (K-NN) — untuk menemukan pola atau hubungan antar atribut data.

6. Interpretation / Evaluation (Interpretasi dan Evaluasi)

Hasil dari proses data mining diinterpretasikan menjadi informasi yang

bermakna. Misalnya, model K-NN dapat menunjukkan seberapa besar kemungkinan seorang mahasiswa lulus tepat waktu.

Evaluasi dilakukan untuk mengukur tingkat akurasi model agar dapat dipercaya sebelum diimplementasikan.

7. Knowledge (Pengetahuan)

Tahap akhir berupa hasil pengetahuan yang diperoleh dari keseluruhan proses. Pengetahuan ini bisa digunakan oleh pihak kampus untuk membuat kebijakan atau perbaikan akademik, seperti memberikan bimbingan

kepada mahasiswa berisiko tidak lulus tepat waktu.

HASIL DAN PEMBAHASAN

1. Analisa

Pembagian data dilakukan dengan proporsi 80% untuk data latih dan 20% untuk data uji. Data latih digunakan untuk membangun model prediksi, sedangkan data uji digunakan untuk mengukur performa model terhadap data baru yang belum pernah dilihat sebelumnya.

Tabel 1. Data Latih (80%)

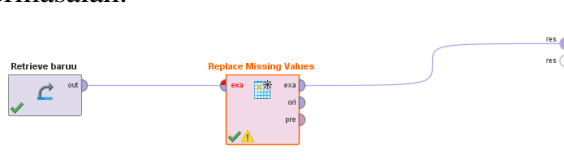
No	Program Studi	Jenis Kelamin	Angkatan	IPK	SKS	Kehadiran (%)	Status Kelulusan
1	Teknik Informatika	Laki-laki	2019	3.72	144	98	Lulus Tepat Waktu
2	Teknik Informatika	Perempuan	2019	3.45	144	96	Lulus Tepat Waktu
3	Sistem Informasi	Laki-laki	2018	3.12	140	92	Tidak Tepat Waktu
4	Teknik Informatika	Perempuan	2019	3.81	144	99	Lulus Tepat Waktu
5	Sistem Informasi	Laki-laki	2018	2.95	138	90	Tidak Tepat Waktu
6	Teknik Informatika	Perempuan	2019	3.55	144	97	Lulus Tepat Waktu
7	Sistem Informasi	Perempuan	2018	3.22	142	94	Lulus Tepat Waktu
8	Teknik Informatika	Laki-laki	2019	3.6	144	95	Lulus Tepat Waktu

Tabel 2. Data Uji (20%)

No	Program Studi	Jenis Kelamin	Angkatan	IPK	SKS	Kehadiran (%)	Status Kelulusan
1	Teknik Informatika	Perempuan	2019	3.4	144	94	Lulus Tepat Waktu
2	Sistem Informasi	Laki-laki	2018	3.05	138	91	Tidak Tepat Waktu
3	Teknik Informatika	Laki-laki	2019	3.78	144	99	Lulus Tepat Waktu
4	Sistem Informasi	Perempuan	2018	3.18	142	93	Tidak Tepat Waktu

2. Proses Pembersihan Data (Data Cleaning)

Pada penelitian ini, tahap pembersihan data dilakukan terlebih dahulu untuk memastikan kualitas data yang baik. Hal ini penting karena data awal mengandung *noise* atau kerusakan yang dapat memengaruhi hasil analisis. Oleh karena itu, dilakukan proses modifikasi, pengisian nilai yang hilang, serta penghapusan data yang tidak valid atau bermasalah.



Gambar 3. Tahap Pembersihan Data (Preprocessing Data)

a. Selection Data

Pemilihan data dilakukan dengan menyeleksi informasi yang relevan dan

menghapus bagian data yang tidak diperlukan

Tabel 3. Seleksi Data

No	Nama Atribut	Gunakan (✓ / ✗)	Alasan Pemilihan / Penghapusan
1	prodi	✗	Hanya satu nilai ("Teknik Industri"), tidak variatif
2	predikat	✓	Mewakili kualitas kelulusan
3	tanggal_lulus	✗	Redundan, sudah tercermin di lama_studi
4	tanggal_masuk	✗	Redundan, sudah tercermin di lama_studi
5	jenis_kelamin	✓	Variabel demografis yang relevan
6	tahun_lahir	✓	Digunakan untuk menghitung usia
7	status_peg	✓	Menunjukkan status pekerjaan mahasiswa

8	ips_1	✓	Indikator performa akademik
9	ips_2	✓	Indikator performa akademik
10	ips_3	✓	Indikator performa akademik
11	ips_4	✓	Indikator performa akademik
12	ips_5	✓	Indikator performa akademik
13	ips_6	✓	Indikator performa akademik
14	ips_7	✓	Indikator performa akademik

15	ips_8	✓	Indikator performa akademik
16	sks_a	✓	Total SKS yang diambil relevan terhadap kelulusan
17	lama_studi	✓	Mewakili kecepatan menyelesaikan studi
18	lulus_tepat_w	✓ (Label)	Target klasifikasi
19	angkatan	✓	Bisa memengaruhi beban kurikulum
20	lulus_umur	✓	Bisa berkorelasi dengan performa

b. Transformasi Data

Tabel 4. Atribut input

No	Nama Atribut	Keterangan	Tipe Data Asli	Fungsi	Perlakuan Transformasi
1	prodi	Program studi mahasiswa	Nominal	Dihapus	Tidak digunakan (hanya satu program)
2	predikat	Predikat kelulusan	Nominal	Input	Gunakan langsung
3	tanggal_lulus	Tanggal mahasiswa lulus	Tanggal	Dihapus	Tidak digunakan langsung
4	tanggal_masuk	Tanggal mahasiswa masuk kuliah	Tanggal	Dihapus	Sudah diwakili oleh lama_studi
5	jenis_kelamin	Laki-laki / Perempuan	Nominal	Input	Gunakan langsung
6	tahun_lahir	Tahun kelahiran	Numerik	Input	Normalisasi
7	status_pegawai	Status pekerjaan	Nominal	Input	Gunakan langsung
8	ips_1	IP Semester 1	Numerik	Input	Normalisasi
9	ips_2	IP Semester 2	Numerik	Input	Normalisasi
10	ips_3	IP Semester 3	Numerik	Input	Normalisasi
11	ips_4	IP Semester 4	Numerik	Input	Normalisasi
12	ips_5	IP Semester 5	Numerik	Input	Normalisasi
13	ips_6	IP Semester 6	Numerik	Input	Normalisasi
14	ips_7	IP Semester 7	Numerik	Input	Normalisasi
15	ips_8	IP Semester 8	Numerik	Input	Normalisasi
16	sks_a	Jumlah total SKS	Numerik	Input	Normalisasi
17	lama_studi	Lama studi dalam hari	Numerik	Input	Normalisasi
18	lulus_tepat_w	Status lulus tepat waktu atau tidak	Nominal	Output	Dijadikan Label
19	angkatan	Tahun angkatan masuk kuliah	Numerik	Input	Bisa dinormalisasi
20	lulus_umur	Umur saat lulus	Numerik	Input	Normalisasi

Tabel 5. Atribut output

No	Nama Atribut	Keterangan	Tipe Data	Fungsi	Perlakuan Transformasi
1	Lulus tepat waktu	Status Kelulusan tepat waktu atau tidak	Nominal	Output (Label)	Dijadikan Label Klasifikasi

c. Pembahasan

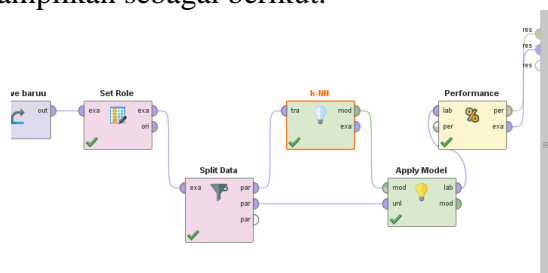
Proses memasukkan data ke dalam RapidMiner dilakukan melalui fitur impor data yang mempunyai 19 atribut, sebagai berikut:

Row No.	prodi	predikat	tanggal_kulu	tanggal_ma...	jenis_kelamin	tahun_lahir	status_paga...	ips_1	ips_2
1	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015	Laki-laki	1994	Bekerja	322333333	3421
2	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015	Penerempuan	1996	Bekerja	330555555	3421
3	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015	Laki-laki	1984	Bekerja	322555555	3324
4	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015	Laki-laki	1996	Bekerja	323444444	3286
5	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015	Laki-laki	1984	Bekerja	314222222	3394
6	Teknik Industri	Pujian	Aug 28, 2019	Sep 6, 2015	Laki-laki	1995	Bekerja	292222222	3252
7	Teknik Industri	Pujian	Aug 28, 2019	Sep 6, 2015	Laki-laki	1982	Bekerja	355777777	3676
8	Teknik Industri	Pujian	Aug 30, 2019	Sep 6, 2015	Laki-laki	1990	Bekerja	305888888	3111
9	Teknik Industri	Sangat Mem...	Aug 31, 2022	Sep 6, 2015	Laki-laki	1989	Bekerja	319555555	3251
10	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015	Laki-laki	1995	Tidak bekerja	323444444	3394
11	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015	Laki-laki	1986	Bekerja	305777777	3386
12	Teknik Industri	Pujian	Aug 28, 2019	Sep 6, 2015	Laki-laki	1995	Tidak bekerja	316888888	3356
13	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015	Laki-laki	1996	Bekerja	294666666	2956
14	Teknik Industri	Sangat Mem...	Jan 28, 2021	Sep 6, 2015	Laki-laki	1996	Tidak bekerja	333555555	3252
15	Teknik Industri	Pujian	Aug 30, 2019	Sep 6, 2015	Laki-laki	1995	Bekerja	319777777	3603
16	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015	Laki-laki	1987	Bekerja	319666666	3465

Gambar 4. Import Data

d. Pengujian data di Rapidminer

Dalam penelitian ini, RapidMiner dimanfaatkan sebagai alat bantu untuk pengolahan data. Data yang dianalisis berasal dari data mahasiswa, Selanjutnya, data tersebut diuji dengan menerapkan algoritma K-Nearest Neighbor (K-NN), dengan tahapan pengolahan ditampilkan sebagai berikut.



Gambar 5. Pengolahan data pada rapidminer menggunakan metode K-NN

e. Confusion Matrix

Confusion matrix merupakan alat evaluasi performa dari proses klasifikasi dengan metode K-NN, yang menyajikan metrik seperti akurasi, presisi, dan recall.

Tabel 5. Performa Kinerja K-NN

Accuracy: 95.88%

	True Tepat Waktu	True Tidak Tepat Waktu	Class Precision
Pred. Tepat Waktu	860	35	96.09%
Pred. Tidak Tepat Waktu	2	0	0.00%
Class Recall	99.77%	0.00%	—

Berdasarkan hasil pengujian model K-Nearest Neighbor (K-NN) menggunakan RapidMiner, diperoleh nilai akurasi sebesar 95.88%. Hal ini menunjukkan bahwa

model mampu memprediksi kelulusan mahasiswa dengan tingkat ketepatan yang sangat tinggi. Dari tabel di atas terlihat bahwa sebagian besar data mahasiswa yang lulus tepat waktu dapat diklasifikasikan dengan benar oleh model, sedangkan kelas 'tidak tepat waktu' belum terprediksi optimal karena jumlah datanya relatif sedikit.

Perhitungan untuk mengukur tingkat akurasi, presisi, dan recall disajikan sebagai berikut:

Rumus Akurasi:

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar \times 100}{TP + TN + FP + FN}$$

Keterangan:

- TP (True Positive) : 860 (pred. Tepat waktu yang benar)
- TN (True Negative) : 0 (pred. tidak tepat waktu yang benar)
- FP (False Positive) : 35 (pred. tepat waktu tapi sebenarnya tidak)
- FN (False Negative) : 2 (pred. tidak tepat waktu tapi sebenarnya tepat)

$$Accuracy = \frac{860 + 0}{860 + 0 + 35 + 2} \times 100\% = \frac{860}{897} \times 100\% = 95.88\%$$

Rumus Precision (Presisi):

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

- TP (True Positive) : 860
- FP (False Positive) : 35

$$Precision = \frac{860}{860 + 35} \times 100\% = \frac{860}{895} \times 100\% = 96.09\%$$

Class Recall

Rumus Recall :

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

- TP (True Positive) : 860
- FN (False Negative): 2

$$Recall = \frac{860}{860 + 2} \times 100\% = \frac{860}{862} \times 100\% = 99.77\%$$

Rumus Precision dan Recall untuk kelas tidak tepat waktu

$$Precision: \frac{TN}{TN + FN} = \frac{0}{0 + 2} = 0\%$$

$$Recall : \frac{TN}{TN + FP} = \frac{0}{0 + 35} = 0\%$$

Model K-Nearest Neighbor (K-NN) menunjukkan akurasi tinggi sebesar 95,88%, dengan precision 96,09% dan recall 99,77% dalam memprediksi

mahasiswa yang meraih kelulusan tepat pada waktunya. Namun, model tidak mampu mengenali mahasiswa yang tidak lulus tepat waktu, ditunjukkan oleh nilai precision dan recall 0% untuk kelompok tersebut. Hal ini menunjukkan bahwa model lebih cenderung mengklasifikasikan semua mahasiswa sebagai lulus tepat waktu, kemungkinan karena mahasiswa yang tidak menyelesaikan studi tepat waktu jumlahnya relatif kecil.

Row No.	lulus_tepat_w	prediction_tu...	confidence...	confidence...	predi	predikat	tanggal_mus	tanggal_ma...
1	Tepat waktu	Tepat waktu	1	0	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
2	Tepat waktu	Tepat waktu	1	0	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
3	Tepat waktu	Tepat waktu	1	0	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
4	Tepat waktu	Tepat waktu	1	0	Teknik Industri	Pujian	Aug 30, 2019	Sep 6, 2015
5	Tepat waktu	Tepat waktu	1	0	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
6	Tidak tepat w	Tepat waktu	0.800	0.200	Teknik Industri	Sangat Mem.	Aug 31, 2022	Sep 6, 2015
7	Tepat waktu	Tepat waktu	0.800	0.200	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
8	Tepat waktu	Tepat waktu	0.800	0.200	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
9	Tepat waktu	Tepat waktu	0.800	0.200	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
10	Tepat waktu	Tepat waktu	1	0	Teknik Industri	Pujian	Aug 28, 2019	Sep 6, 2015
11	Tepat waktu	Tepat waktu	0.800	0.200	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
12	Tepat waktu	Tepat waktu	1	0	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
13	Tepat waktu	Tepat waktu	0.800	0.200	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
14	Tepat waktu	Tepat waktu	1	0	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
15	Tepat waktu	Tepat waktu	0.800	0.200	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015
16	Tepat waktu	Tepat waktu	1	0	Teknik Industri	Pujian	Aug 31, 2019	Sep 6, 2015

Gambar 6. Hasil Prediki Algoritma K-NN

SIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Nearest Neighbor (K-NN) yang diterapkan menggunakan RapidMiner dapat digunakan secara efektif dalam upaya memprediksi kelulusan mahasiswa, algoritma ini menunjukkan akurasi sebesar 95,88%, precision 96,09%, dan recall 99,77% pada kategori mahasiswa yang lulus tepat waktu, model ini menunjukkan kinerja yang baik dalam mengklasifikasikan mayoritas mahasiswa yang sesuai target kelulusan. Namun, kinerja model masih kurang optimal dalam mendeteksi mahasiswa yang mengalami keterlambatan kelulusan, seperti ditunjukkan oleh nilai precision dan recall sebesar 0% pada kategori tersebut. Hal ini menunjukkan bahwa meskipun K-NN memiliki performa tinggi dalam kasus mayoritas, perlu dilakukan perbaikan pada distribusi data atau pemilihan algoritma tambahan agar dapat lebih seimbang dalam menangani seluruh kategori prediksi.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada semua pihak yang telah memberikan dukungan, serta kepada

sumber-sumber penelitian yang menjadi referensi penting dalam penyusunan karya ini.

DAFTAR PUSTAKA

Rani, S. R., Andani, S. R., & Suhendro, D. (2019). *Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Kelulusan Siswa pada SMK Anak Bangsa*. *September*, 670–676.

Saputra, A. Y., & Primadasa, Y. (2018). Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Techno.Com*, 17(4), 395–403.
<https://doi.org/10.33633/tc.v17i4.1864>

Tafonao, I. P. ., & Sibero*, A. F. K. (2022). Teknik Klasifikasi Prediksi Kelulusan Mahasiswa Sistem Informasi Universitas Sari Mutiara Indonesia Menggunakan K- Nearest Neighbors. *Jurnal Mahajana Informasi*, 7(1), 83–90.
<https://doi.org/10.51544/jurnalmi.v7i1.2956>

Susanto, E. S., Kusriani, K., & Al Fatta, H. (2018). *Prediksi kelulusan mahasiswa magister teknik informatika Universitas Amikom Yogyakarta menggunakan metode K-Nearest Neighbor*. *Jurnal Teknologi Informasi Respati (JTIR)*, 13(2), 67–72.

Rani, S. R., Andani, S. R., & Suhendro, D. (2019). *Penerapan algoritma K-Nearest Neighbor untuk prediksi kelulusan siswa pada SMK Anak Bangsa*. *Prosiding Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA)*, 670–676.

Nikmatun, I. A., Waspada, I., & Diponegoro, U. (2019). *Implementasi data mining untuk klasifikasi masa studi mahasiswa menggunakan algoritma K-Nearest Neighbor*. *Jurnal Sistem Informasi (JSI)*, 10(2), 421–432.

Suryadi, L., Ngajiyanto, N., Pratiwi, N. E., Ardhy, F., & Riswanto, P. (2022). *Penerapan data mining prediksi*

- penjualan mebel terlaris menggunakan metode K-Nearest Neighbor (K-NN) (Studi kasus: Toko Zerita Meubel).* Jurnal Sistem Informasi Musirawas (JUSIM), 7(2), 174–184.
- Ahmed, A. A., & Khan, M. A. (2020). *A predictive model for student performance using K-Nearest Neighbor algorithm.* International Journal of Advanced Computer Science and Applications (IJACSA), 11(5), 92–99.
- Kumar, V., & Singh, D. (2021). *Machine learning-based prediction of student academic performance using KNN and decision tree algorithms.* Education and Information Technologies, 26(4), 4567–4582.
- Rahman, M. M., & Islam, M. R. (2019). *Predicting student performance using K-Nearest Neighbor (KNN) algorithm.* International Journal of Data Science and Advanced Analytics, 5(3), 101–110.