

Implementasi Algoritma K-Nearest Neighbor Untuk Menentukan Klasifikasi Kelulusan Mahasiswa Teknik Informatika

Suswantia Andriani, Alwis Nazir*, Reski Mai Candra, Fadhilah Syafria, Iis Afrianty

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹11950125226@students.uin-suska.ac.id, ^{2*}alwis.nazir@uin-suska.ac.id, ³reski.candra@uin-suska.ac.id,

⁴fadhilah.syafria@uin-suska.ac.id, ⁵iis.afrianty@uin-suska.ac.id

Correspondence Author Email: alwis.nazir@uin-suska.ac.id

Submitted: 23/07/2023; Accepted: 25/08/2023; Published: 29/08/2023

Abstrak—Kinerja sebuah universitas dapat dinilai semakin baik jika terdapat peningkatan jumlah mahasiswa yang berhasil menyelesaikan masa studinya tepat waktu. Perguruan tinggi harus meningkatkan kualitas akademik mahasiswa dalam proses perkuliahan untuk hasil yang optimal. Namun kenyataannya banyak keterlambatan kelulusan mahasiswa pada Universitas Islam Negeri Sultan Syaif Kasim Riau terkhusus jurusan Teknik Informatika ini yang menjadi permasalahan. Oleh karena itu dibutuhkan sebuah analisa mengklasifikasi data kelulusan mahasiswa dengan memanfaatkan proses data mining. Penelitian ini menerapkan algoritma *K-Nearest Neighbor* (K-NN) pada proses klasifikasi. Tujuan penelitian ini adalah hasil dari klasifikasi kelulusan ini diharapkan dapat memberikan kontribusi bagi pihak fakultas teknik dan universitas dalam melakukan evaluasi dan perbaikan terhadap sistem pembelajaran, sehingga menghasilkan lulusan tepat waktu dan berkualitas. Data yang digunakan pada penelitian ini sebanyak 613 data 5 tahun terakhir dari tahun 2016 hingga 2020 dengan pembagian data dengan rasio 80 data untuk pelatihan (training) dan 20 data untuk pengujian (testing). Hasil evaluasi *confusion matrix* dari $K = 3$, $K = 5$, $K = 7$ menghasilkan akurasi tertinggi diperoleh ketika $K = 3$ dengan akurasi 93,06%, presisi 99,09%, dan recall 99,58%. Dari hasil penelitian, dapat ditarik kesimpulan bahwa penerapan data mining berhasil menciptakan model klasifikasi dengan memanfaatkan algoritma *K-Nearest Neighbor* (K-NN) dalam mengklasifikasikan status kelulusan mahasiswa pada program sudi Teknik Informatika di Universitas Islam Negeri Sultan Syarif Kasim Riau

Kata Kunci: Data Mining; Kelulusan Mahasiswa; Klasifikasi; *K-Nearest Neighbor*; Universitas

Abstract—The performance of a university can be assessed as getting better if there is an increase in the number of students who successfully complete their studies on time. Universities must improve academic the quality of students lecture process for optimal results. However, in reality there are many delays in student graduation at the Sultan Syaif Kasim Riau State Islamic University, especially the Informatics Engineering major, which is a problem. Therefore an analysis is needed to classify student graduation data by utilizing the data mining process. This study uses the *K-Nearest Neighbor* (K-NN) algorithm in classification process. The purpose of this study is that the results of this graduation classification are expected to assist engineering faculties and universities in evaluating and improving the learning system so that as a result it can produce timely and quality graduates. The data used in this study is 613 data for the last 5 years from 2016 to 2020 with a data distribution ratio of 80 for training data and 20 data for testing. The result of evaluation of the confuion matrix from $K = 3$, $K = 5$, $K = 7$, produce the highest accuracy obtained when $k = 3$, with 93.06% accuracy, 99.09% precision and 99.58% recall. From the result of the study, it can be concluded that the application of data mining has succeeded in creating a classification model by utilizing the *K-Nearest Neighbor* (K-NN) algorithm for the classification of Informatics Engineering students at Sultan Syarif Kasim Riau State Islamic University.

Keywords: Classification; Data Mining; *K-Nearest Neighbor*; Student Graudation; University

1. PENDAHULUAN

Pendidikan adalah suatu aktivitas yang melibatkan semua individu di dunia, tanpa memandang tempat dan waktu, karena pendidikan dapat ditemukan di manapun dan kapanpun di dunia [1]. Pendidikan pada hakikatnya mempunyai tujuan untuk menggali potensi peserta didik (mahasiswa) sehingga mereka menjadi individu yang memiliki iman dan taqwa kepada Tuhan Yang Maha Esa, berprilaku baik, waran, berpengetahuan, berkompeten, mandiri, serta menjadi warga negara yang demokratis dan bertanggung jawab [2]. Lembaga pendidikan adalah universitas yang menyampaikan pengajaran akademik kepada mahasiswa [3]. Peran mahasiswa sangat signifikan dalam mencapai kesuksesan sebuah perguruan tinggi, baik itu swasta maupun negeri [4][5].

Perguruan tinggi memiliki keharusan untuk menyelenggarakan pendidikan yang unggul bagi para mahasiswa agar menghasilkan manusia yang terampil, berpengetahuan luas, dan mampu bersaing di dunia kerja [6][7][8]. Mahasiswa memiliki peranan yang sangat krusial dalam suatu perguruan tinggi sebagai aset berharga bagi institusi pendidikan [6]. Oleh karena itu, penting untuk memberikan perhatian pada tingkat kelulusan mahasiswa agar bisa menyelesaikan studi tepat waktu sesuai jadwal yang ditetapkan [9]. Kinerja sebuah universitas dapat dinilai sebagai semakin baik jika terdapat peningkatan jumlah mahasiswa yang berhasil menyelesaikan studinya tepat waktu [10]. Oleh karena itu diperlukan pemantauan dan evaluasi terhadap kecenderungan mahasiswa dalam mencapai kelulusan tepat waktu atau tidak.

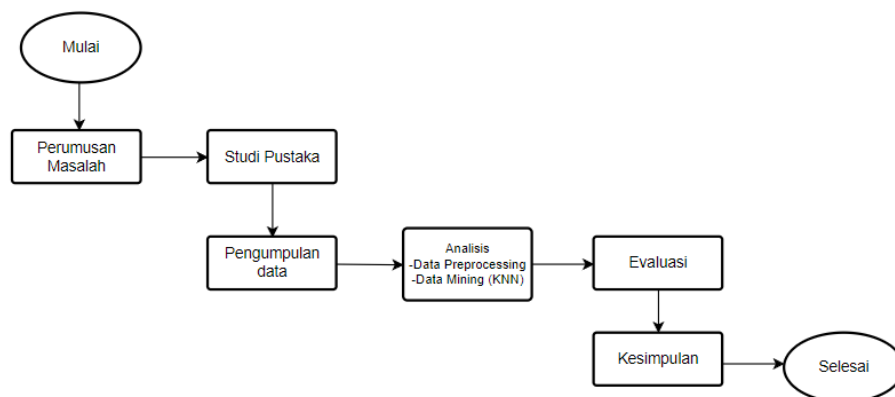
Keterlambatan kelulusan mahasiswa ini menjadi dilema di setiap perguruan tinggi. Saat ini Universitas Islam Negeri Sultan Syarif Kasim Riau juga menghadapi tantangan terkait keterlambatan kelulusan mahasiswa, terutama di jurusan Teknik Informatika. Berdasarkan data lima tahun belakangan, tingkat kelulusan mahasiswa dalam jurusan ini tergolong rendah, yang tentunya merupakan masalah serius bagi jurusan Teknik Informatika sehingga menyebabkan lambatnya untuk jurusan ini naik akreditasi A. Oleh karena itu, diperlukan analisis untuk

mengklasifikasikan data kelulusan mahasiswa jurusan Teknik Informatika di Universitas tersebut. Dengan hasil klasifikasi kelulusan ini, pihak fakultas teknik dan universitas akan mendapatkan dukungan untuk melakukan evaluasi dan perbaikan sistem pembelajaran. Tujuannya adalah agar dapat menghasilkan lulusan yang tepat waktu dan berkualitas tinggi. Masalah tersebut bisa diatasi dengan cara membuat suatu model Data Mining dengan algoritma K-Nearest Neighbor (K-NN). Proses pembelajaran data mining melibatkan pengembalian dan identifikasi data, yang selanjutnya diolah menjadi pengetahuan dan wawasan yang bernilai [11][12]. K-Nearest Neighbor (K-NN) merupakan suatu algoritma yang digunakan untuk mencari kasus dengan melakukan estimasi sejauh mana kemiripan antara kasus baru dengan kasus yang sudah ada sebelumnya [13][14].

Beberapa penelitian yang telah berhasil dilakukan terkait penerapan metode klasifikasi menerapkan algoritma K-Nearest Neighbor (K-NN), seperti yang terlihat pada penelitian [15] dalam upaya memprediksi kelulusan mahasiswa, digunakan tiga pendekatan yaitu naïve bayes, K-Nearest Neighbor (K-NN), dan Neural Network. Dari beberapa algoritma ini K-Nearest Neighbor (K-NN) menjadi yang paling baik dengan nilai akurasi 89%. Penelitian [16] berdasarkan kinerja bentuk klasifikasi dari algoritma yang digunakan dalam studi ini, dapat disimpulkan bahwa model K-Nearest Neighbor (K-NN) menghasilkan kinerja yang unggul dalam nilai matriks dan evaluasi, jika dibandingkan dengan algoritma Decision Tree dan Random Forest. Pada penelitian [3] nilai akurasi dari algoritma K-Nearest Neighbor (K-NN) lebih baik dibanding dengan Naïve Bayes. K-Nearest Neighbor (K-NN) mendapatkan akurasi 100% dan Naïve Bayes 99,29%. Sedangkan penelitian [17] hasil dari penggunaan data mining menyatakan pengklasifikasian masa studi yang optimal dapat dicapai dengan memilih parameter dari seluruh mata kuliah, mencapai akurasi sebesar 75,95%. Dan pada penelitian [18] didapatkan nilai akurasi metode K-Nearest Neighbor (K-NN) dengan akurasi 70,28%, error 29,78%, dan auc sebesar 0,739.

2. METODOLOGI PENELITIAN

Tahapan penelitian untuk menentukan klasifikasi kelulusan mahasiswa Teknik Informatika Universitas Negeri Sultan Syarif Kasim Riau dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

Pada gambar 1 penulis akan menjelaskan tentang tahapan-tahapan penelitian untuk mengklasifikasikan kelulusan mahasiswa Teknik Informatika Universitas Negeri Sultan Syarif Kasim Riau.

2.1 Perumusan Masalah

Tahapan awal yang sangat penting dalam metodologi penelitian adalah merumuskan masalah. Penelitian yang akan dilakukan bertujuan untuk mengidentifikasi permasalahan yang ada dan kemudian menemukan solusi yang diharapkan. Rumusan masalah utamanya adalah bagaimana menerapkan metode *K-Nearest Neighbor* (K-NN) dalam mengklasifikasi kelulusan mahasiswa di jurusan teknik informatika Uin Suksa Riau, serta mengukur tingkat akurasi.

2.2 Studi Pustaka

Dalam teknik pengumpulan data dalam sebuah penelitian, langkah pertama yang biasanya dilakukan adalah studi pustaka, yang sering disebut juga sebagai *literature review*. Metode ini melibatkan pencarian sumber referensi dari berbagai buku, karya ilmiah, dan jurnal-jurnal yang dianggap *relevan* dan mendukung.

2.3 Pengumpulan Data

Pada proses pengumpulan data, dilakukan dengan mengumpulkan data kelulusan mahasiswa program studi Teknik Informatika. Pengumpulan data yang dilaksanakan adalah dengan memperoleh data sekunder, yaitu data yang diperoleh melalui sumber buku, jurnal dan penelitian sebelumnya. Data yang diperoleh berasal dari Pusat

Teknologi Informasi dan Pangkalan Data (PTIPD) Universitas Islam Negeri Sultan Syarif Kasim Riau, yang mencakup informasi mengenai kelulusan mahasiswa 2016-2020 sebanyak 613 data kelulusan terdiri dari Jenis Kelamin, Indeks Prestasi kumulatif, Asal Sekolah, Jalur Masuk, Nilai Sistem Informasi (SI), Nilai Jaringan Komputer (Jarkom), dan Nilai Rekayasa Perangkat Lunak Berorientasi Objek (RPLBO), data tersebut nantinya akan digunakan sebagai pembentuk model yang nantinya akan diteliti.

2.4 Analysis

2.4.1 Data Preprocessing

Pada tahap ini, data akan mengalami transformasi melalui penerapan metode pengelompokan berdasarkan label. Selain itu, proses *transformasi* juga dilakukan dengan menerapkan normalisasi *min-max*, sebelum melakukan klasifikasi, data perlu melewati tahap *preprocessing*, yang mencakup langkah-langkah sebagai berikut :

a. Data Selection

Setelah proses pengumpulan data selesai, peneliti kemudian melanjutkan ke tahap data *selection*. Data *selection* merupakan proses pemilihan atribut-atribut tertentu dari kumpulan data yang lebih besar. Dalam penelitian ini, penulis memilih atribut-atribut yang berpengaruh terhadap kelulusan mahasiswa Teknik Informatika, yaitu Jenis Kelamin, Indeks Prestasi kumulatif, Asal Sekolah, Jalur Masuk, Nilai Sistem Informasi (SI), Nilai Jaringan Komputer (Jarkom), dan Nilai Rekayasa Perangkat Lunak Berorientasi Objek (RPLBO).

b. Transformasi Data

Dikarenakan metode *K-Nearest Neighbor* (K-NN) hanya dapat digunakan pada data yang berupa angka, maka poses *transformasi* data bertujuan untuk mengubah atribut-atribut *non-numerik* menjadi nilai angka, sehingga data tersebut dapat diolah menerapkan metode *K-nearest Neighbor* (K-NN).

Tabel 1. Transformasi Jenis Kelamin

No	Jenis Kelamin	Keterangan
1	P	1
2	L	2

Pada tabel 1 penulis melakukan *Transformasi* data dari atribut nominal berubah menjadi numeric seperti Jenis kelamin P berubah menjadi 1 , dan Jenis Kelamin L berubah menjadi 2.

Tabel 2. Transformasi Jalur Masuk

No	Jalur Masuk	Keterangan
1	PBUD	1
2	SNMPTN	2
3	SBMPTN	3
4	UMUM	4

Pada tabel 2 penulis melakukan *Transformasi* data dari atribut nominal berubah menjadi numeric seperti Jalur Masuk PBUD berubah menjadi 1 , Jalur Masuk SNMPTN berubah menjadi 2, Jalur Masuk SBMPTN berubah menjadi 3, dan Jalur Masuk UMUM berubah menjadi 4.

Tabel 3. Transformasi Asal Sekolah

No	Asal Sekolah	Keterangan
1	MAN	1
2	SMK	2
3	SMA	3

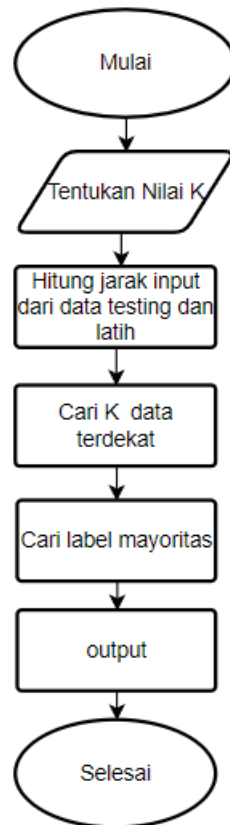
Pada tabel 3 penulis melakukan *Transformasi* data dari atribut nominal berubah menjadi numeric seperti Asal Sekolah MAN berubah menjadi 1 , Asal Sekolah SMK berubah menjadi 2, dan Asal Sekolah SMA berubah menjadi 3.

Proses *transformasi* juga melibatkan normalisasi *min-max*, dimana data diubah menjadi rentang antara 0 dan 1. Metode ini mempertahankan korelasi dengan data asli dan dapat diaplikasikan menggunakan rumus yang telah ditetapkan sebelumnya.

$$\text{normalized}(x) = \frac{\text{minRange} + (x - \text{minValue})(\text{maxRange} - \text{minRange})}{\text{maxValue} - \text{minValue}} \quad (1)$$

2.5 Data Mining

Setelah data dikumpulkan dan diproses dalam tahap pengumpulan data dan *preprocessing* data, selanjutnya akan dilakukan pengolahan menerapkan algoritma *K-Nearest Neighbor* (K-NN). Detail penggunaan algoritma *K-Nearest Neighbor* (K-NN) dapat dilihat pada *flowchart* yang terdapat pada Gambar 2 dibawah ini.



Gambar 2. Flowchart Algoritma *K-Nearest Neighbor* (K-NN)

Berdasarkan Gambar 2 diatas Penjelasan berikut ini menggambarkan bagaimana proses perhitungan dilakukan menggunakan *algoritma K-Nearest Neighbor* (K-NN).

- Pada proses pertama, nilai k diputuskan/ditetapkan.
- Melakukan perhitungan jarak Euclidean data testing dan data uji.

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

Keterangan

$d(x, y)$ = Jarak data x ke y

X_i = data testing ke-1

Y_i = data training ke-1

- Golongkan data menggunakan perhitungan jarak Euclidean.
- Membentuk kluster tetangga terdekat dengan mengurutkan data berdasarkan jarak terkecil.
- Memilih label yang paling sering muncul di antara K tetangga dan menggunakan sebagai hasil prediksi.

2.6 Evaluasi

Evaluasi kinerja algoritma *K-Nearest Neighbor* (K-NN) memakai 7 parameter yaitu Jenis Kelamin, Indeks Prestasi kumulatif, Asal Sekolah, Jalur Masuk, Nilai Sistem Informasi (SI), Nilai Jaringan Komputer (Jarkom), dan Nilai Rekayasa Perangkat Lunak Berorientasi Objek (RPLBO). Tahapan evaluasi menggunakan *Confusion Matrix* yang dapat diterapkan menggunakan perangkat lunak Rapidminer. *Confusion Matrix* digunakan untuk mengukur tingkat akurasi, presisi, dan recall.

2.7 Kesimpulan

Kesimpulan dapat ditarik dari hasil perhitungan menerapkan metode *K-Nearest Neighbor* (K-NN) yang menghasilkan nilai-nilai akurasi, presisi, dan *recall*.

3. HASIL DAN PEMBAHASAN

Pada analisis ini, data yang digunakan berasal dari Pusat Teknologi Informasi dan Pangkalan Data (PTIPD) Universitas Islam Negeri Sultan Syarif Kasim Riau. Berikut adalah tabel kelulusan mahasiswa Teknik Informatika dapat dilihat pada tabel 4 dibawah ini.

Tabel 4. Dataset Kelulusan mahasiswa

No	Jenis kelamin	IPK	Asal Sekolah	Jalur Masuk	Nilai_SI	Nilai_Jar kom	Nilai_RPLBO	Semester
1	L	2,75	MAN	SBMPTN	3,00	2,00	2,30	Tidak Tepat Waktu
2	L	2,54	MAN	PBUD	3,30	2,70	2,30	Tidak Tepat Waktu
3	L	2,59	MAN	SBMPTN	3,30	2,30	2,30	Tidak Tepat Waktu
4	L	2,75	SMK	SBMPTN	3,70	2,30	3,30	Tidak Tepat Waktu
5	L	2,61	SMA	SBMPTN	3,70	3,20	3,00	Tidak Tepat Waktu
6	L	3,29	SMA	SNMPTN	4,00	3,70	3,00	Sedang
7	L	2,81	SMA	SNMPTN	4,00	4,00	4,00	Tidak Tepat Waktu
8	L	2,80	SMA	SNMPTN	4,00	3,0	3,00	Tidak Tepat Waktu
9	P	3,28	SMA	SNMPTN	3,70	3,30	3,70	Sedang
10	L	3,14	SMA	SBMPTN	3,70	2,70	2,70	Tidak Tepat Waktu
	--	--	--	--	--	--	--	--
	--	--	--	--	--	--	--	--
612	P	3,64	MAN	SNMPTN	2,70	2,30	3,00	Sedang
613	L	3,54	SMK	PBUD	3,00	2,70	3,50	Sedang

Pada Tabel 4 diatas ini merupakan tabel dataset kelulusan mahasiswa teknik informatika, Sebelum menerapkan algoritma *K-Nearest Neighbor* (K-NN) untuk perhitungan, peneliti melaksanakan langkah *preprocessing* atau *transformasi* data sesuai dengan konteks permasalahan yang ada, dengan mengadaptasi data agar sesuai dengan algoritma yang dipergunakan. Data yang berhasil melalui tahap *transformasi* data kemudian ditampilkan dalam tabel 5.

Tabel 5. Dataset Kelulusan mahasiswa

No	Jenis kelamin	IPK	Asal Sekolah	Jalur Masuk	Nilai_SI	Nilai_Jar kom	Nilai_RPLBO	Semester
1	2	2,75	1	3	3,00	2,00	2,30	Tidak Tepat Waktu
2	2	2,54	1	1	3,30	2,70	2,30	Tidak Tepat Waktu
3	2	2,59	1	3	3,30	2,30	2,30	Tidak Tepat Waktu
4	2	2,75	2	3	3,70	2,30	3,30	Tidak Tepat Waktu
5	2	2,61	3	3	3,70	3,20	3,00	Tidak Tepat Waktu
6	2	3,29	3	2	4,00	3,70	3,00	Sedang
7	2	2,81	3	2	4,00	4,00	4,00	Tidak Tepat Waktu
8	2	2,80	3	2	4,00	3,0	3,00	Tidak Tepat Waktu
9	1	3,28	3	2	3,70	3,30	3,70	Sedang
10	2	3,14	3	3	3,70	2,70	2,70	Tidak Tepat Waktu
	--	--	--	--	--	--	--	--
	--	--	--	--	--	--	--	--
612	1	3,64	1	2	2,70	2,30	3,00	Sedang
613	2	3,54	2	1	3,00	2,70	3,50	Sedang

Berdasarkan tabel 5 diatas ini merupakan tabel dataset kelulusan mahasiswa teknik informatika Universitas Sultan Syarif Kasim Riau yang telah di *preprocessing* pada atribut Jenis Kelamin, Asal Sekolah, Jalur Masuk.

3.1 Pengolahan Data

Sistem yang dimanfaatkan untuk melakukan klasifikasi status kelulusan mahasiswa dalam jurusan Teknik Informatika melibatkan penggunaan data numerik dan non-numerik. Setelah data terkumpul, penulis akan menerapkan algoritma *K-Nearest Neighbor* (K-NN) untuk mengolah data tersebut yang menciptakan hasil akurasi, presisi, dan *recall*. Berikut langkah-langkah dalam memproses data menerapkan algoritma *K-Nearest Neighbor* (K-NN).

- Data dibagi dalam rasio 80 data untuk pelatihan (*training*) dan 20 data untuk pengujian (*testing*), yang diperlihatkan dalam Tabel 6 sebagai data *training* dan Tabel 7 sebagai data *testing*.

Tabel 6. Dataset Training

No	Jenis kelamin	IPK	Asal Sekolah	Jalur Masuk	Nilai_SI	Nilai_Jar kom	Nilai_RPLBO	Semester
1	2	2,75	1	3	3,00	2,00	2,30	Tidak Tepat Waktu
2	2	2,54	1	1	3,30	2,70	2,30	Tidak Tepat Waktu
3	2	2,59	1	3	3,30	2,30	2,30	Tidak Tepat Waktu
4	2	2,75	2	3	3,70	2,30	3,30	Tidak Tepat Waktu

No	Jenis kelamin	IPK	Asal Sekolah	Jalur Masuk	Nilai_SI	Nilai_Jar kom	Nilai_RPLBO	Semester
5	2	2,61	3	3	3,70	3,20	3,00	Tidak Tepat Waktu
6	2	3,29	3	2	4,00	3,70	3,00	Sedang
7	2	2,81	3	2	4,00	4,00	4,00	Tidak Tepat Waktu
8	2	2,80	3	2	4,00	3,0	3,00	Tidak Tepat Waktu
9	1	3,28	3	2	3,70	3,30	3,70	Sedang
10	2	3,14	3	3	3,70	2,70	2,70	Tidak Tepat Waktu
--	--	--	--	--	--	--	--	--
--	--	--	--	--	--	--	--	--
489	1	3,24	2	3	3,70	1,70	2,30	Tidak Tepat Waktu
490	2	3,66	3	3	3,00	3,30	3,70	Sedang

Berdasarkan tabel 6 diatas merupakan dataset training yang mana pembagian data set dengan rasio 80:20. Data training diatas dari data no 1 – 490.

Tabel 7. Dataset *Testing*

No	Jenis kelamin	IPK	Asal Sekolah	Jalur Masuk	Nilai_SI	Nilai_Jar kom	Nilai_RPLBO	Semester
491	2	3,30	2	2	4,00	3,30	3,30	Tidak Tepat Waktu
492	2	2,77	3	3	3,70	2,70	4,00	Tidak Tepat Waktu
493	1	3,22	1	3	3,00	2,70	2,70	Sedang
494	2	3,08	3	2	3,70	3,00	4,00	Tidak Tepat Waktu
495	1	3,68	3	2	3,00	3,00	3,30	Sedang
496	2	2,89	2	3	3,70	3,70	3,30	Tidak Tepat Waktu
497	1	3,31	3	4	2,70	3,00	3,30	Sedang
498	1	3,25	1	3	3,70	2,70	2,30	Sedang
499	1	3,27	2	3	3,00	4,00	2,30	Sedang
450	1	3,26	3	2	3,30	1,70	2,30	Sedang
--	--	--	--	--	--	--	--	--
--	--	--	--	--	--	--	--	--
612	1	3,64	1	2	2,70	2,30	3,00	Sedang
613	2	3,54	2	1	3,00	2,70	3,50	Sedang

Berdasarkan tabel 6 diatas merupakan dataset testing yang mana pembagian data set dengan rasio 80:20. Data testing diatas dari data no 490 - 613.

- b. Hasil normalisasi min-max dari data training dan testing ditampilkan dalam tabel dibawah ini.

Tabel 8. Dataset *Training*

No	Jenis kelamin	IPK	Asal Sekolah	Jalur Masuk	Nilai_SI	Nilai_Jar kom	Nilai_RPLBO	Semester
1	1	0,17	0	0,66	0,66	0,33	0,43	Tidak Tepat Waktu
2	1	0	0	0	0,76	0,56	0,43	Tidak Tepat Waktu
3	1	0,04	0	0,66	0,76	0,43	0,43	Tidak Tepat Waktu
4	1	0,17	0,5	0,66	0,9	0,43	0,76	Tidak Tepat Waktu
5	1	0,05	1	0,66	0,9	0,74	0,66	Tidak Tepat Waktu
6	1	0,62	1	0,33	1	0,9	0,66	Sedang
7	1	0,22	1	0,33	1	1	1	Tidak Tepat Waktu
8	1	0,21	1	0,33	1	0,66	0,66	Tidak Tepat Waktu
9	0	0,61	1	0,33	0,9	0,76	0,9	Sedang
10	1	0,5	1	0,66	0,9	0,56	0,56	Tidak Tepat Waktu
--	--	--	--	--	--	--	--	--
--	--	--	--	--	--	--	--	--
489	0	0,58	0,5	0,66	0,9	0,23	0,43	Tidak Tepat Waktu
490	1	0,93	1	0,66	0,66	0,76	0,9	Sedang

Berdasarkan tabel 8 diatas dilakukannya proses normalisasi min-max sehingga bisa mendapatkan hasil akurasi yang lebih akurat .[19] pada dataset training kelulusan mahasiswa tekni informatika.

Tabel 9. Dataset *Testing*

No	Jenis kelamin	IPK	Asal Sekolah	Jalur Masuk	Nilai_SI	Nilai_Jar kom	Nilai_RPLBO	Semester
491	1	0,52	0,35	0,33	1	0,74	0,68	Tidak Tepat Waktu
492	1	0	0,71	0,66	0,88	0,52	1	Tidak Tepat Waktu
493	0	0,44	0	0,66	0,60	0,52	0,40	Sedang
494	1	0,30	0,71	0,33	0,88	0,63	1	Tidak Tepat Waktu
495	0	0,90	0,71	0,33	0,60	0,63	0,68	Sedang
496	1	0,11	0,35	0,66	0,88	0,88	0,68	Tidak Tepat Waktu
497	0	0,53	0,71	1	0,49	0,63	0,68	Sedang
498	0	0,47	0	0,66	0,88	0,52	0,22	Sedang
499	0	0,49	0,35	0,66	0,60	1	0,22	Sedang
450	0	0,48	0,71	0,33	0,72	0,15	0,22	Sedang
	--	--	--	--	--	--	--	--
	--	--	--	--	--	--	--	--
612	0	0,86	0	0,33	0,49	0,37	0,54	Sedang
613	1	0,76	0,35	0	0,60	0,52	0,77	Sedang

Berdasarkan tabel 9 diatas dilakukannya proses normalisasi min-max sehingga bisa mendapatkan hasil akurasi yang lebih akurat .[20] pada dataset testing kelulusan mahasiswa tekni informatika

- c. Penerapan metode euclidean distance digunakan untuk menghitung jarak antara data *training* dan data *testing*. Berikut adalah contoh perhitungan jarak antara data *testing* 491 dengan data *training* 1 adalah .

$$\text{Distance} = \sqrt{(1-1)^2 + (0,17-0,52)^2 + (0-0,35)^2 + (0,66-0,33)^2 + (0,66-1)^2 + (0,33-0,74)^2 + (0,43-0,68)^2} = 0,83858471$$

Selanjutnya adalah contoh perhitungan jarak antara data latih 491 dengan data uji 2 adalah.

$$\text{Distance} = \sqrt{(1-1)^2 + (0-0,52)^2 + (0-0,35)^2 + (0-0,33)^2 + (0,76-1)^2 + (0,56-0,74)^2 + (0,43-0,68)^2} = 0,814264959$$

Selanjutnya adalah contoh perhitungan jarak antara data latih 491 dengan data uji 3 adalah.

$$\text{Distance} = \sqrt{(1-1)^2 + (0,04-0,52)^2 + (0-0,35)^2 + (0,66-0,33)^2 + (0,76-1)^2 + (0,43-0,74)^2 + (0,43-0,68)^2} = 0,828094224$$

Setelah selesai menghitung jarak dari data latih 491 ke beberapa data uji. Lakukanlah perhitungan jarak dari data *testing* berikutnya dengan menerapkan proses perhitungan yang sama seperti pada data sebelumnya.

3.2 Evaluasi

Setelah menyelesaikan proses pemodelan *K-Nearest Neighbor* (K-NN), dilakukan pengujian akurasi untuk mengevaluasi performa model klasifikasi *K-Nearest Neighbor* (K-NN) yang telah dimodelkan sebelumnya. Akurasi diukur menggunakan *confusion matrix* melalui penggunaan alat Rapidminer. Hasil dari *confusion matrix* tersebut dipresentasikan dalam bentuk gambar dengan menampilkan beberapa percobaan dari nilai k.

accuracy: 93.06%

	true Tidak Tepat Waktu	true Sedang	true Tepat Waktu	class precision
pred. Tidak Tepat Waktu	237	32	0	88.10%
pred. Sedang	1	217	1	99.09%
pred. Tepat Waktu	0	0	2	100.00%
class recall	99.58%	87.15%	66.67%	

Gambar 3. Evaluasi K =3

accuracy: 81.43%

	true Tidak Tepat Waktu	true Sedang	true Tepat Waktu	class precision
pred. Tidak Tepat Waktu	196	46	1	80.66%
pred. Sedang	42	203	2	82.19%
pred. Tepat Waktu	0	0	0	0.00%
class recall	82.35%	81.53%	0.00%	

Gambar 4. Evaluasi K =5

accuracy: 80.82%

	true Tidak Tepat Waktu	true Sedang	true Tepat Waktu	class precision
pred. Tidak Tepat Waktu	197	50	1	79.44%
pred. Sedang	41	199	2	82.23%
pred. Tepat Waktu	0	0	0	0.00%
class recall	82.77%	79.92%	0.00%	

Gambar 5. Evaluasi K =7

Berdasarkan hasil dari seluruh percobaan yang telah dilaksanakan dari $K = 3$, $K = 5$, $K = 7$, diketahui bahwa klasifikasi yang terbaik tercapai ketika menggunakan nilai $k = 3$. Pada kondisi ini, diperoleh akurasi sebesar 93,06%, presisi 99,09% dan recall 99,58%. Dari hasil percobaan tersebut, dapat disimpulkan bahwa penggunaan nilai k dalam proses klasifikasi berpengaruh pada tingkat akurasi.

4. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa penerapan data mining telah berhasil menghasilkan model klasifikasi dengan menerapkan algoritma *K-Nearest Neighbor* (K-NN) dalam mengklasifikasikan status kelulusan mahasiswa pada program studi Teknik Informatika di Universitas Islam Negeri Sultan Syarif Kasim Riau. Terdapat variasi akurasi hasil pengujian yang dilakukan dari $k = 3$, $k = 5$, dan $k = 7$ menunjukkan evaluasi $K = 3$ memiliki hasil tertinggi dengan akurasi 93,06%, presisi 99,09%, dan recall 99,58%. Dari eksperimen yang telah dilakukan, terungkap bahwa nilai k , dan tahapan *preprocessing* data seperti *transformasi* data dan normalisasi *min-max* yang digunakan memiliki pengaruh pada hasil akhir yang diperoleh. Untuk penelitian lebih lanjut, disarankan untuk menerapkan pendekatan normalisasi data yang berbeda dan mencoba rasio data selain 80 data testing dan 20 data uji.

REFERENCES

- [1] Muhammad Hasan, Tuti Khairani Harahap, Inanna, Uswatun Khasanah, Badroh Rif'ati, A A Musyaffa, Susanti, Sitti Hajerah Hasyim, Nuraisyiah, Ahmad Fuadi, Muh. Suranto, Fakhurrizi, Nur Arisah, Ahmad Zaki, Cahya Edi Setyawan, *Ebook Landasan Pendidikan*. 2021. [Online]. Available: [http://repository.uinjambi.ac.id/10882/1/EBOOK LANDASAN PENDIDIKAN %281%29.pdf#page=50](http://repository.uinjambi.ac.id/10882/1/EBOOK%20LANDASAN%20PENDIDIKAN%20%281%29.pdf#page=50)
- [2] Z. S. Imelia Rosa Br Sinuhaji, "IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR UNTUK MENENTUKAN MAHASISWA BERPRESTASI DI STMIK KRISTEN NEUMANN INDONESIA," *CAHAYATECH*, vol. 3, no. 11, p. 140, 2018, doi: 10.47047/ct.v6i2.10.
- [3] R. Situmorang *et al.*, "MODEL ALGORITMA K-NEAREST NEIGHBOR (K-NN) DAN NAÏVE BAYES UNTUK PREDIKSI KELULUSAN MAHASISWA," vol. 7, no. 1, pp. 250–254, 2023.
- [4] A. Putri, C. S. Hardiana, E. Novfuja, and ..., "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," ... *Indones. J. ...*, vol. 3, no. April, pp. 20–26, 2023, [Online]. Available: <https://journal.irpi.or.id/index.php/malcom/article/view/610>
- [5] R. Muliono, J. H. Lubis, and N. Khairina, "Analisis Algoritma K-Nearest Neighbors dalam Prediksi Waktu Kelulusan Mahasiswa," *Semant. (Seminar Nas. Tek. Inform.)*, vol. 2, no. 1, pp. 12–16, 2019.
- [6] I. P. J. Tafonao, A. F. K. Sibero, and D. Mining, "Sistem Informasi Universitas Sari Mutiara Indonesia Menggunakan K- Nearest Neighbors," vol. 7, no. 1, pp. 83–90, 2022.
- [7] M. F. Abadi, "KLASIFIKASI LAMA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," *J. Teknol. Pint.*, vol. 2, no. 12, pp. 1–17, 2022, [Online]. Available: <http://teknologipintar.org/index.php/teknologipintar/article/view/247%0Ahttp://teknologipintar.org/index.php/teknologipintar/article/download/247/234>
- [8] A. J. Nathan and A. Scobell, "Model Algoritma K-nearest Neighbor untuk memprediksi kelulusan mahasiswa," *Foreign Aff.*, vol. 91, no. 5, pp. 1–9, 2022.
- [9] L. Anshori, R. R. M. Putri, and Tibyani, "Implementasi Metode K-Nearest Neighbor untuk Rekomendasi Keminatan Studi (Studi Kasus : Jurusan Teknik Informatika Universitas Brawijaya)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 7, pp. 2745–2753, 2018, [Online]. Available: https://www.researchgate.net/publication/321097020_Implementasi_Metode_K-Nearest_Neighbor_untuk_Rekomendasi_Keminatan_Studi_Studi_Kasus_Jurusan_Teknik_Informatika_Universitas_Brawijaya
- [10] O. S. D. Putri and Y. P. P. Anggi, "Prediksi Kelulusan Mahasiswa dengan Metode K-Nearest Neighbor pada Jurusan Sistem Informasi Institut XYZ," *Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform.*, pp. 199–206, 2023.
- [11] B. Basuki, A. Nazir, S. K. Gusti, L. Handayani, and I. Iskandar, "Klasifikasi Tingkat Keberhasilan Produksi Ayam Broiler di Riau Menggunakan Algoritma K-Nearest Neighbor," *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, p. 510, 2023, doi: 10.30865/json.v4i3.5800.

- [12] N. Nursobah, S. Lailiyah, B. Harpad, and M. Fahmi, "Penerapan Data Mining Untuk Prediksi Perkiraan Hujan dengan Menggunakan Algoritma K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, 2022, doi: 10.47065/bits.v4i3.2564.
- [13] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, p. 133, 2021, doi: 10.21107/nero.v6i2.237.
- [14] D. Prasetyawan and R. Gatra, "Algoritma K-Nearest Neighbor untuk Memprediksi Prestasi Mahasiswa Berdasarkan Latar Belakang Pendidikan dan Ekonomi," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 56–67, 2022, doi: 10.14421/jiska.2022.7.1.56-67.
- [15] D. Safitri, S. S. Hilabi, and F. Nurapriani, "Analisis Penggunaan Algoritma Klasifikasi Dalam Prediksi Kelulusan Menggunakan Orange Data Mining," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 1, pp. 75–81, 2023, doi: 10.36341/rabit.v8i1.3009.
- [16] LOUIS MADAERDO SOTARJUA and DIAN BUDHI SANTOSO, "Perbandingan Algoritma Knn, Decision Tree, Dan Random Forest Pada Data Imbalanced Class Untuk Klasifikasi Promosi Karyawan," *J. INSTEK (Informatika Sains dan Teknol.*, vol. 7, no. 2, pp. 192–200, 2022, doi: 10.24252/instek.v7i2.31385.
- [17] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
- [18] Manarul Hidayat*; Ahmad Faqih; Tati Suprpti, "IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOUR UNTUK PREDIKSI KETEPATAN KELULUSAN," vol. 10, no. 2, 2022.
- [19] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor," *J. Inform. MULTI*, vol. 1, no. 3, pp. 200–214, 2023, [Online]. Available: <https://jurnal.publikasitecno.id/index.php/jim203>
- [20] R. Iriane, "KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Data Mining Untuk Prediksi Penjualan Produk Pangan Hewan Menggunakan Metode K-Nearest Neighbor," *Media Online*, vol. 3, no. 5, pp. 509–515, 2023, [Online]. Available: <https://djournals.com/klik>