

# MODEL ALGORITMA KNN UNTUK PREDIKSI KELULUSAN MAHASISWA STIKOM CKI

<sup>1</sup>Tiara Ratu Alifia, <sup>2</sup>Tundo, <sup>3\*</sup>Muhammad Syazidan, <sup>4</sup>Faldo Satria  
<sup>1,2,3\*,5</sup>Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika  
(STIKOM CKI)  
<sup>1</sup>tiararatu42@gmail.com, <sup>2</sup>asna8mujahid@gmail.com, <sup>3\*</sup>syazidan313@gmail.com,  
<sup>4</sup>faldosatria12@gmail.com  
) Penulis Korespondensi

## Abstrak

Penelitian ini mengembangkan model prediksi kelulusan mahasiswa menggunakan algoritma K-Nearest Neighbor (KNN) dengan mempertimbangkan variabel usia, Indeks Prestasi Kumulatif (IPK), jumlah Satuan Kredit Semester (SKS), partisipasi dalam tes TOEFL, kegiatan seminar, dan partisipasi dalam Kuliah Kerja Praktek (KKP). Data dari 80 mahasiswa program studi teknik informatika dan sistem informasi di STIKOM Cipta Karya Informatika dianalisis untuk melatih dan menguji model. Hasilnya menunjukkan bahwa model KNN dengan  $K=3$ ,  $K=4$ ,  $K=5$  menghasilkan akurasi prediksi sebesar 66,67%. Faktor IPK dan jumlah SKS memiliki pengaruh signifikan terhadap kelulusan, sementara partisipasi dalam KKP dan tes TOEFL juga berkontribusi. Sertifikat seminar dan usia memiliki pengaruh yang lebih rendah. Temuan ini mengindikasikan bahwa algoritma KNN efektif untuk prediksi kelulusan mahasiswa, memberikan wawasan bagi institusi pendidikan untuk meningkatkan program akademik dan pembinaan mahasiswa.

**Kata Kunci:** Algoritma KNN, prediksi kelulusan, IPK, SKS, sertifikat seminar, KKP.

## Abstract

This study develops a student graduation prediction model using the K-Nearest Neighbor (KNN) algorithm, considering variables such as age, Grade Point Average (GPA), number of Credits Earned (CE), participation in TOEFL tests, seminar activities, and participation in internships. Data from 80 students in the computer engineering and information systems programs at STIKOM Cipta Karya Informatika were analyzed to train and test the model. The results show that the KNN model with  $K=3$ ,  $K=4$ , and  $K=5$  produces a prediction accuracy of 66,67%. GPA and the number of credits earned significantly influence graduation, while participation in internships and TOEFL tests also contribute. Seminar certificates and age have a lower impact. These findings indicate that the KNN algorithm is effective for predicting student graduation, providing insights for educational institutions to enhance academic programs and student development.

**Keywords:** KNN Algorithm, graduation prediction, GPA, credits earned, seminar certificates, internsh.

## PENDAHULUAN

Dalam dunia pendidikan tinggi, tingkat kelulusan mahasiswa merupakan salah satu indikator penting dari kualitas pendidikan dan

efektivitas institusi. Tingkat kelulusan yang tinggi mencerminkan keberhasilan program akademik dan pengajaran, sementara tingkat kelulusan yang rendah dapat menunjukkan berbagai masalah, mulai dari kesulitan akademik mahasiswa, ketidakcocokan

program studi, hingga faktor-faktor eksternal seperti masalah finansial dan kesehatan mental[1]. Untuk meningkatkan tingkat kelulusan, banyak institusi pendidikan telah berusaha mengidentifikasi faktor-faktor yang mempengaruhi kesuksesan akademik mahasiswa[2][3]. Salah satu pendekatan yang semakin populer adalah penggunaan teknik data mining dan pembelajaran mesin (*machine learning*) untuk menganalisis data mahasiswa dan memprediksi kemungkinan kelulusan mereka[4][5]. Melalui pendekatan ini, institusi dapat mengambil langkah-langkah proaktif untuk membantu mahasiswa yang berisiko tidak lulus.

Salah satu algoritma yang efektif untuk prediksi ini adalah K-Nearest Neighbors (KNN). KNN adalah algoritma sederhana namun powerful yang digunakan dalam klasifikasi dan regresi[6][7]. Algoritma ini bekerja dengan cara mengidentifikasi 'tetangga' terdekat dari sebuah data baru berdasarkan metrik jarak tertentu, dan kemudian menentukan kelas dari data baru tersebut berdasarkan mayoritas kelas dari tetangga-tetangga terdekatnya[8].

Keunggulan algoritma KNN dalam konteks prediksi kelulusan mahasiswa terletak pada kemampuannya untuk menangani data yang non-linear dan tidak memerlukan asumsi kuat tentang distribusi data[9]. Selain itu, KNN dapat dengan mudah diimplementasikan dan diinterpretasikan, sehingga memudahkan institusi pendidikan dalam memahami faktor-

faktor yang berkontribusi pada kesuksesan atau kegagalan akademik mahasiswa[10].

## **METODE PENELITIAN**

Metode yang digunakan dalam penelitian ini mencakup pengumpulan data yang bersumber dari data akademik dan non-akademik mahasiswa, serta data kelulusan mahasiswa[11]. Data akan diperoleh dari institusi pendidikan yang relevan dan diperiksa untuk keakuratan dan keterwakilan.

Dalam penelitian ini pengumpulan data dilakukan melalui beberapa tahapan. Data akan diperiksa dan disiapkan untuk analisis dengan cara menghapus data yang tidak relevan, menggabungkan data yang berbeda, dan mengkonversi data menjadi format yang sesuai untuk analisis[12].

### **Algoritma K-Nearest Neighbor (K-NN)**

K-Nearest Neighbor, yang sering disingkat K-NN, adalah algoritma yang bertujuan untuk menemukan kesamaan antara data baru dan data lama[13][14]. Algoritma ini bekerja dengan mengelompokkan data baru ke dalam kategori yang paling mirip dengan data yang sudah ada. Dengan kata lain, K-Nearest Neighbor menyimpan semua data lama dan mengklasifikasikan data baru berdasarkan kemiripan. K-Nearest Neighbor (K-NN) adalah salah satu algoritma yang sering digunakan untuk klasifikasi data. Algoritma ini menggunakan jarak terdekat setiap objek untuk melakukan klasifikasi data[15]. Dalam

prosesnya, ada dua jenis data yang terlibat: data uji dan dataset. Langkah selanjutnya melibatkan perhitungan jarak antara data uji dan dataset menggunakan metode Euclidean (*Query Distance*). Rumus *Euclidean (Query Distance)* dapat dilihat Persamaan 1.

*Euclidean Distance*=

$$\sqrt{\sum_k^d = \{XData\ uji^2 - XData\ Latih^2\}} \quad (1)$$

X: Merupakan data point atau vektor yang berisi nilai-nilai dari fitur yang digunakan untuk menghitung jarak.

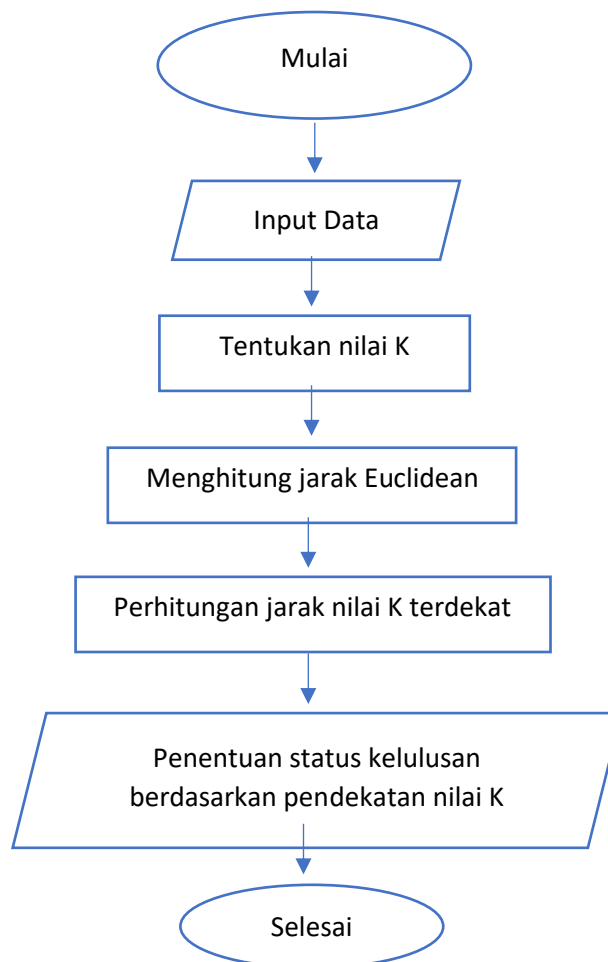
d: Tidak disebutkan dalam dokumen, namun biasanya d adalah jarak yang dihitung.

k: Adalah jumlah tetangga terdekat yang digunakan dalam algoritma KNN.

Persamaan 1 digunakan untuk menghitung jarak antara dua titik dalam ruang fitur, yang kemudian digunakan untuk menentukan tetangga terdekat dalam KNN.

### Tahapan Penelitian

Perancangan sistem adalah menggambarkan sistem yang akan dilakukan dalam prediksi kelulusan mahasiswa STIKOM CKI. Adapun rancangan pada penelitian ini digambarkan dalam bentuk flowchart yang tampak pada gambar 1



**Gambar 1 Flowchart Tahapan Penelitian**

Dengan penjelasan sebagai berikut:

Mulai (Start):

Tahapan ini merupakan langkah awal dari proses penelitian.

Input Data:

Data mahasiswa yang terdiri dari berbagai parameter seperti usia, IPK, jumlah SKS, kelengkapan berkas (sertifikat TOEFL dan seminar), serta partisipasi dalam KKP dimasukkan ke dalam sistem. Data ini kemudian dibagi menjadi dua kelompok, yaitu data latih (80 data) dan data uji (6 data).

Tentukan Nilai K:

Menentukan nilai K yang akan digunakan dalam algoritma K-NN. Nilai K adalah jumlah tetangga terdekat yang akan diperhitungkan dalam klasifikasi.

Menghitung Jarak Euclidean:

Menggunakan persamaan jarak Euclidean untuk menghitung jarak antara data uji dengan setiap data latih. Persamaan jarak Euclidean didefinisikan persamaan 2.

$$\sqrt{\sum_k^d = \{XData\ uji^2 - XData\ Latih^2\}}$$

(2)

Perhitungan Jarak Nilai K Terdekat:

Setelah jarak Euclidean dihitung, langkah selanjutnya adalah mengidentifikasi K data latih yang memiliki jarak terdekat dengan data uji.

Penentuan Status Kelulusan Berdasarkan Pendekatan Nilai K:

Status kelulusan mahasiswa diprediksi berdasarkan mayoritas status kelulusan dari K

tetangga terdekat yang telah diidentifikasi pada langkah sebelumnya.

Selesai (End):

Tahap akhir dari proses penelitian di mana hasil prediksi kelulusan ditentukan dan disimpulkan.

## HASIL DAN PEMBAHASAN

### Analisis Data

Pada tahap ini, data yang digunakan adalah data sekunder yang terdiri dari 86 data. Data tersebut diperoleh dari mahasiswa STIKOM CKI dan telah dibagi menjadi dua kelompok, yaitu 80 data untuk data latih dan 6 data untuk data uji. Pembagian ini berdasarkan praktik umum dalam machine learning, yaitu membagi data menjadi bagian latih dan uji untuk mengevaluasi model. Referensi yang dapat digunakan untuk mendukung tentang pembelajaran mesin yang menyarankan sekitar 70-80% data untuk pelatihan dan sisanya untuk pengujian. Data yang diperoleh mencakup mahasiswa dari semester 8. Seluruh data telah diolah dalam format Excel untuk memudahkan proses pengolahan lebih lanjut. Parameter yang diukur meliputi Usia, IPK, Jumlah SKS, Kelengkapan Berkas (Sertifikat TOEFL dan Sertifikat Seminar), dan tanda bukti ikut serta dalam KKP. Untuk mendapatkan parameter-parameter tersebut, penulis meminta data akademik mahasiswa semester 8 dari manajemen kampus dan membagikan link Google Form yang berisi pertanyaan kepada mahasiswa.

1. Apakah anda mengikuti kegiatan (Ya/Tidak). Data Kontinu: pengambilan sertifikat TOEFL? Merupakan data yang dapat diukur dan memiliki nilai dalam rentang tertentu, misalnya IPK atau jumlah SKS.
2. Apakah anda mengikuti kegiatan seminar dan mendapatkan sertifikat? Dengan menerapkan skor, seperti pada parameter partisipasi dalam kegiatan KKP diberi skor, Ya = 1 poin, Tidak = 0 poin. Begitupun untuk parameter partisipasi dalam pengambilan sertifikat TOEFL dan sertifikat seminar yang diberi persentase, seperti yang mengikuti kedua kegiatan = 100 poin, hanya satu kegiatan = 50, tidak ikut kegiatan sama sekali = 0 poin.

### Konversi Data

Setelah memperoleh data, langkah berikutnya adalah mengonversi data tersebut. Ada beberapa pada Tabel 1, yang awalnya bersifat diskrit diubah menjadi data kontinu. Data Diskrit: Merupakan data yang dapat dihitung dan memiliki nilai terbatas, misalnya jumlah sertifikat atau keikutsertaan dalam

Dengan menerapkan skor, seperti pada parameter partisipasi dalam kegiatan KKP diberi skor, Ya = 1 poin, Tidak = 0 poin. Begitupun untuk parameter partisipasi dalam pengambilan sertifikat TOEFL dan sertifikat seminar yang diberi persentase, seperti yang mengikuti kedua kegiatan = 100 poin, hanya satu kegiatan = 50, tidak ikut kegiatan sama sekali = 0 poin.

**Tabel 1. Data Latih**

N O.	Nama Mahasiswa	Program Studi	Us ia	IP K	Jumlah SKS	Kelengkapan Berkas		Sudah Ikut KKP
						TOE FL	Sertifikat Seminar	
1	AisYah	Teknik Informatika	21	3,80	130	Ya	Tidak	Ya
2	Budi	Teknik Informatika	22	3,48	140	Tidak	Ada	Tidak
3	Chandra	Sistem Informasi	23	3,20	120	Ya	Tidak	Ya
4	Dewi	Teknik Informatika	21	3,59	135	Tidak	Tidak	Ya
5	Erwin	Sistem Informasi	24	3,87	110	Ya	Ada	Tidak
6	Fani	Sistem Informasi	22	3,52	145	Ya	Tidak	Ya
7	Gina	Sistem Informasi	21	3,99	125	Ya	Tidak	Ya
8	Hasan	Teknik Informatika	23	2,30	115	Tidak	Ada	Tidak
9	Indah	Sistem Informasi	21	2,51	138	Ya	Ada	Ya
10	Joko	Teknik Informatika	22	2,73	105	Tidak	Tidak	Tidak
..								
80	Hanif	Teknik Informatika	20	2,90	105	Ya	Tidak	Tidak

Proses ini bertujuan untuk menghadirkan data dalam bentuk yang lebih terstruktur dan dapat diolah lebih lanjut. Setelah proses konversi data, hasilnya dapat dilihat pada Tabel 2. Tabel ini menyajikan representasi baru dari data latih setelah melalui proses skoring. Konversi data diskrit menjadi data kontinu dilakukan untuk memungkinkan penggunaan algoritma yang membutuhkan data dalam bentuk numerik kontinu. Misalnya, keikutsertaan dalam kegiatan KKP yang

awalnya berupa Ya/Tidak diubah menjadi 1/0 untuk memudahkan perhitungan dan analisis lebih lanjut terkait prediksi kelulusan mahasiswa menggunakan algoritma K-NN.

### Klasifikasi Data Latih

Mempermudah proses klasifikasi maka nama parameter dirubah menjadi simbol  $x$ , seperti  $x_1 = \text{Usia}$ ,  $x_2 = \text{IPK}$ ,  $x_3 = \text{Jumlah SKS}$ ,  $x_4 = \text{Kelengkapan Berkas}$ ,  $x_5 = \text{Sudah Ikut KKP}$ , dan dapat dilihat pada Tabel 3.

**Tabel 2. Hasil Konversi Data Latih**

<b>N O.</b>	<b>Nama Mahasiswa</b>	<b>Program Studi</b>	<b>Usia</b>	<b>IPK</b>	<b>Jumlah SKS</b>	<b>Kelengkapan Berkas</b>	<b>Sudah Ikut KKP</b>
1	AisYah	Teknik Informatika	21	3,80	130	50	1
2	Budi	Teknik Informatika	22	3,48	140	50	1
3	Chandra	Sistem Informasi	23	3,20	120	50	0
4	Dewi	Teknik Informatika	21	3,59	135	0	1
5	Erwin	Sistem Informasi	24	3,87	110	100	0
6	Fani	Sistem Informasi	22	3,52	145	50	1
7	Gina	Sistem Informasi	21	3,99	125	50	1
8	Hasan	Teknik Informatika	23	2,30	115	50	0
9	Indah	Sistem Informasi	21	2,51	138	100	1
10	Joko	Teknik Informatika	22	2,73	105	0	0
80	Hanif	Teknik Informatika	20	2,90	105	50	0

**Tabel 3. Hasil Perubahan Simbol**

NO.	Nama Mahasiswa	Program Studi	Parameter				
			x1	x2	x3	x4	x5
1	AisYah	Teknik Informatika	21	3,80	130	50	1
2	Budi	Teknik Informatika	22	3,48	140	50	1
3	Chandra	Sistem Informasi	23	3,20	120	50	0
4	Dewi	Teknik Informatika	21	3,59	135	0	1
5	Erwin	Sistem Informasi	24	3,87	110	100	0
6	Fani	Sistem Informasi	22	3,52	145	50	1
7	Gina	Sistem Informasi	21	3,99	125	50	1
8	Hasan	Teknik Informatika	23	2,30	115	50	0
9	Indah	Sistem Informasi	21	2,51	138	100	1
10	Joko	Teknik Informatika	22	2,73	105	0	0
80	Hanif	Teknik Informatika	20	2,90	105	50	0

**Tabel 4. Data Uji**

N O.	Nama Mahasiswa	Program Studi	Us ia	IP K	Jumlah SKS	Kelengkapan Berkas	Sudah Ikut KKP
1	Raden	Teknik Informatika	22	3,30	145	50	1
2	Bagas	Sistem Informasi	21	3,00	105	50	0
3	Pramudya	Sistem Informasi	24	3,38	115	50	1
4	Syafira	Teknik Informatika	22	3,45	138	50	1
5	Aulia	Sistem Informasi	21	3,23	125	100	0
6	Syakila	Teknik Informatika	23	3,48	130	100	1

Berikut adalah 6 buah data yang dipilih untuk diujikan dengan menggunakan algoritma KNN dengan menggunakan K=3, K=4, K=5 yang terlihat pada Tabel 4.

**Penyelesaian Algoritma K-Nearest**

**Neighbor**

Langkah – langkah menghitung prediksi kelulusan mahasiswa STIKOM CKI menggunakan Algoritma K-Nearest Neighbor yaitu :

1. Nilai K yang digunakan adalah 3,4, dan 5.

2. Menghitung jarak Euclid (*Query Distance*) masing-masing data latih dengan data uji menggunakan rumus:

*Euclidean Distance* =

$$\sqrt{\sum_k^d = \{XData\ uji^2 - XData\ Latih^2\}}$$

Hasil dari perhitungan *Query Distance* data latih dengan data uji terdapat pada Tabel 5.

Selanjutnya mengurutkan dari hasil perhitungan *Query Distance* data latih dengan data uji berdasarkan jarak terkecil dan menentukan data yang masuk ke K=3, K=4,

K=5 yang sudah ditentukan di awal. Data tersebut dapat dilihat pada Tabel 6. Selanjutnya yaitu pengumpulan hasil dari perhitungan data di atas yang termasuk pada klasifikasi Y=Kategori Nearest Neighbor dengan nilai K=3, K=4, K=5. Tabel ini mengurutkan hasil perhitungan jarak terkecil dan menentukan status kelulusan berdasarkan nilai K=3. Dapat dilihat pada Tabel 7. Tabel ini mengurutkan hasil perhitungan jarak terkecil dan menentukan status kelulusan berdasarkan nilai K=4. Dapat dilihat pada Tabel 8

**Tabel 5. Hasil Perhitungan *Query Distance* Data Latih Dengan Data Uji**

N	Data Uji 1	Data Uji 2	Data Uji 3	Data Uji 4	Data Uji 5	Data Uji 6
0	15,041608	25,032778	15,302823		50,262559	50,041007
1	96	51	27	8,0698513	62	18
2	5,0032389	35,031848	25,080071	2,0002249	52,221283	
	51	37	77	87	98	51
3	25,040167	15,134067	5,1992691	18,057200	50,289172	51,000768
	73	53	79	78	79	62
4	51,000824	58,321077	53,935555	50,100095	100,50437	100,14495
	5	66	06	81	6	54
5	61,076385	50,346369	50,261716	57,351341	15,310440	20,053730
	78	28	05	75	88	33
6		40,028369	30,066918	7,0003499	53,870994	52,211125
	0,22	94	7	91	98	25
7	20,036868	20,049441	10,458111	13,049582	50,015773	50,291749
	52	39	68	37	51	82
8	30,049958	10,222035	1,7794381	23,072115	51,037877	52,224442
	4	02	14	2	11	55
9	50,503703	59,918612	55,124920	50,018832	13,058269	8,3030656
	82	3	86	45	41	99
10	64,041587	50,010727	51,043339	59,920934	101,98651	103,09006
	27	85	43	57	87	98
..	..	..	..	..	..	..
80	40,064448	1,0049875	10,827298	33,080243	53,861942	55,994074
	08	62	83	35	97	69

**Tabel 6. Pengurutan Jarak Terkecil dari *Query Distance***

Data Uji 1	Data Uji 2	Data Uji 3
------------	------------	------------



Euclid Distance	Jarak Terkecil	Status	Euclid Distance	Jarak Terkecil	Status	Euclid Distance	Jarak Terkecil	Status
0,15	1	Lulus	1,005	1	Lulus	1,779	1	Lulus
0,22	2	Lulus	2,002	2	Lulus	2,241	2	Lulus
1,073	3	Tidak Lulus	3,002	3	Tidak Lulus	2,248	3	Tidak Lulus
1,114	4	Lulus	5,39	4	Lulus	3,176	4	Lulus
2,036	5	Tidak Lulus	5,393	5	Tidak Lulus	3,237	5	Tidak Lulus
5,003	6		10	6		3,355	6	
5,018	7		10,005	7		5,197	7	
5,102	8		10,027	8		5,199	8	
5,102	9		10,051	9		5,215	9	
5,104	10		10,053	10		5,231	10	
Data Uji 4			Data Uji 5			Data Uji 6		
Euclid Distance	Jarak Terkecil	Status	Euclid Distance	Jarak Terkecil	Status	Euclid Distance	Jarak Terkecil	Status
0,37	1	Lulus	1,532	1	Lulus	1,19	1	Lulus
0,49	2	Lulus	3,351	2	Lulus	2,352	2	Lulus
0,76	3	Tidak Lulus	3,447	3	Tidak Lulus	2,533	3	Tidak Lulus
1,68	4	Lulus	5	4	Lulus	5,103	4	Lulus
2	5	Tidak Lulus	5,557	5	Tidak Lulus	5,105	5	Tidak Lulus
2,08	6		10,061	6		5,168	6	
2,239	7		10,069	7		5,434	7	
2,259	8		10,088	8		5,549	8	
2,26	9		10,1	9		8,074	9	
2,266	10		10,109	10		8,303	10	

**Tabel 7. Hasil Kualifikasi Y=Kategori Nearest Neighbor K=3**

NO	Nama	Program Studi	Lulu	Tidak	Status
----	------	---------------	------	-------	--------

NO	Mahasiswa	Program Studi	Lulus	Tidak Lulus	Status Kelulusan
1	Raden	Teknik Informatika	3	0	Lulus
2	Bagas	Sistem Informasi	0	3	Tidak Lulus
3	Pramudya	Sistem Informasi	0	3	Tidak Lulus
4	Syafira	Teknik Informatika	3	0	Lulus
5	Aulia	Sistem Informasi	2	1	Lulus
6	Syakila	Teknik Informatika	1	2	Tidak Lulus

**Tabel 8. Hasil Kualifikasi Y=Kategori Nearest Neighbor K=4**

NO	Nama Mahasiswa	Program Studi	Lulus	Tidak Lulus	Status Kelulusan
1	Raden	Teknik Informatika	4	0	Lulus
2	Bagas	Sistem Informasi	0	4	Tidak Lulus
3	Pramudya	Sistem Informasi	0	4	Tidak Lulus
4	Syafira	Teknik Informatika	3	1	Lulus
5	Aulia	Sistem Informasi	2	2	Tidak lulus
6	Syakila	Teknik Informatika	2	2	Lulus

**Tabel 9. Hasil Kualifikasi Y=Kategori Nearest Neighbor K=5**

NO	Nama Mahasiswa	Program Studi	Lulus	Tidak Lulus	Status Kelulusan
1	Raden	Teknik Informatika	5	0	Lulus
2	Bagas	Sistem Informasi	0	5	Tidak Lulus
3	Pramudya	Sistem Informasi	0	5	Tidak Lulus
4	Syafira	Teknik Informatika	4	1	Lulus
5	Aulia	Sistem Informasi	2	3	Tidak Lulus
6	Syakila	Teknik Informatika	3	2	Lulus

Tabel ini mengurutkan hasil perhitungan jarak terkecil dan menentukan status kelulusan berdasarkan nilai K=5. Dapat dilihat pada Tabel 9

#### **Pengujian Confussion Matrix**

*Confusion matrix* adalah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi dalam *machine learning*[16]. Tabel ini menunjukkan jumlah prediksi yang benar dan salah yang dibuat oleh model

dengan membandingkan prediksi model dengan nilai sebenarnya. *Confusion matrix* terdiri dari empat komponen utama, yaitu:

1. **True Positive (TP)**: Jumlah data yang benar-benar positif dan diprediksi positif oleh model.
2. **True Negative (TN)**: Jumlah data yang benar-benar negatif dan diprediksi negatif oleh model.
3. **False Positive (FP)**: Jumlah data yang sebenarnya negatif tetapi diprediksi positif oleh model (juga dikenal sebagai *Type I Error*).
4. **False Negative (FN)**: Jumlah data yang sebenarnya positif tetapi

diprediksi negatif oleh model (juga dikenal sebagai *Type II Error*).

5. Berikut adalah hasil dari pengujian data *Actual* dan prediksi dengan *Confussion Matrix*, pada tabel 10 sampai 15.

$$Accuracy : \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision : \frac{TP}{TP+FP} \quad (4)$$

$$Recall : \frac{TP}{TP+FN} \quad (5)$$

**Tabel 10. Perbandingan Data Actual dan Prediksi K=3**

Lulus/ tidak lulus (Instansi)	Class Prediksi
lulus	lulus
lulus	Tidak lulus
Tidak lulus	Tidak lulus
lulus	lulus
lulus	lulus
lulus	Tidak lulus

**Tabel 11. Perbandingan Data Actual dan Prediksi K=4**

Lulus/ tidak lulus (Instansi)	Class Prediksi
Lulus	lulus
lulus	Tidak lulus
lulus	Tidak lulus
Lulus	lulus
Tidak lulus	Tidak lulus
lulus	Lulus

**Tabel 12. Perbandingan Data Actual dan Prediksi K=5**

Lulus/ tidak lulus (Instansi)	Class Prediksi
lulus	lulus
Tidak lulus	Tidak lulus
lulus	Tidak lulus
lulus	lulus
lulus	Tidak lulus
Lulus	lulus

**Tabel 13. Pengujian Performa K=3 dengan *Confussion Matrix***

Actual	Predicted		Total
	Lulus	Tidak lulus	
Lulus	TP = 3	FN = 1	4
Tidak lulus	FP = 1	TN = 1	2
	4	2	6

**Tabel 14. Pengujian Performa K=4 dengan Confussion Matrix**

Actual	Predicted		Total
	Lulus	Tidak lulus	
Lulus	TP = 3	FN = 2	5
Tidak lulus	FP = 0	TN = 1	1
	3	3	6

**Tabel 15. Pengujian Performa K=5 dengan Confussion Matrix**

Actual	Predicted		Total
	Lulus	Tidak lulus	
Lulus	TP = 3	FN = 2	5
Tidak lulus	FP = 0	TN = 1	1
	3	3	6

**Accuracy :**

$$\frac{TP + TN}{TP + TN + FP + FN} = \frac{3 + 1}{2 + 1 + 2 + 1} = \frac{4}{6} \times 100\% = 66,67\%$$

**Precision :**

$$\frac{TP}{TP + FP} = \frac{3}{3 + 1} = \frac{3}{4} \times 100\% = 75\%$$

**Recall :**

$$\frac{TP}{TP + FN} = \frac{3}{4} \times 100\% = 75\%$$

**Accuracy :**

$$\frac{TP + TN}{TP + TN + FP + FN} = \frac{3 + 1}{3 + 1 + 0 + 2} = \frac{4}{6} \times 100\% = 66,67\%$$

**Precision :**

$$\frac{TP}{TP + FP} = \frac{3}{3 + 0} = \frac{3}{3} \times 100\% = 100\%$$

**Recall :**

$$\frac{TP}{TP + FN} = \frac{3}{3 + 2} = \frac{3}{5} \times 100\% = 60\%$$

**Accuracy :**

$$\frac{TP + TN}{TP + TN + FP + FN} = \frac{3 + 1}{3 + 1 + 0 + 2} = \frac{4}{6} \times 100\% = 66,67\%$$

**Precision :**

$$\frac{TP}{TP + FP} = \frac{3}{3 + 0} = \frac{3}{3} \times 100\% = 100\%$$

**Recall :**

$$\frac{TP}{TP + FN} = \frac{3}{3 + 2} = \frac{3}{5} \times 100\% = 60\%$$

Berdasarkan perhitungan di atas, dapat disimpulkan bahwa pengujian hasil algoritma K-NN dengan data asli dan data prediksi menunjukkan hasil pada K=3, dengan Accuracy 66,67%, Precision 75%, dan Recall 75%. Sebagai perbandingan, untuk K=4 dan K=5, nilai Accuracy sama-sama 66,67%, dengan Precision 100% dan Recall 60%. Hasil ini digunakan untuk memprediksi kelulusan mahasiswa di STIKOM CKI Jakarta.

## KESIMPULAN DAN SARAN

Penelitian ini menghasilkan model prediksi kelulusan mahasiswa menggunakan algoritma K-Nearest Neighbor (KNN), dengan mempertimbangkan variabel-variabel

seperti usia, IPK, jumlah SKS, partisipasi dalam tes TOEFL, kegiatan seminar, dan magang (KKP). Data dari 80 mahasiswa program studi Teknik Informatika dan Sistem Informasi di STIKOM Cipta Karya Informatika digunakan untuk melatih dan menguji model ini.

Hasilnya menunjukkan bahwa model KNN dengan  $K=3$ ,  $K=4$ , dan  $K=5$  mencapai akurasi prediksi sebesar 66,67%. IPK dan jumlah SKS terbukti memiliki pengaruh signifikan terhadap kelulusan mahasiswa, sementara partisipasi dalam KKP dan tes TOEFL juga memberikan kontribusi positif. Sebaliknya, sertifikat seminar dan usia mahasiswa memiliki pengaruh yang lebih rendah.

Kesimpulannya, algoritma KNN efektif dalam memprediksi kelulusan mahasiswa, Model KNN yang diimplementasikan dalam penelitian ini menunjukkan akurasi sebesar 66,67%. Meskipun akurasi ini tidak sangat tinggi, tetap menunjukkan bahwa model memiliki kemampuan untuk memprediksi kelulusan mahasiswa lebih baik daripada tebakan acak. Penelitian menunjukkan bahwa IPK dan jumlah SKS merupakan faktor yang signifikan dalam memprediksi kelulusan. Ini menunjukkan bahwa model KNN dapat mengidentifikasi variabel penting yang mempengaruhi kelulusan. Partisipasi dalam KKP dan tes TOEFL juga berkontribusi pada prediksi kelulusan, meskipun dengan pengaruh yang lebih rendah dibandingkan

IPK dan jumlah SKS. Faktor-faktor ini menunjukkan bahwa model dapat mempertimbangkan berbagai aspek akademik dan non-akademik.

Meskipun akurasinya tidak sangat tinggi, model ini memberikan wawasan yang dapat digunakan institusi pendidikan untuk meningkatkan program akademik dan pembinaan mahasiswa.

Dengan demikian, meskipun akurasi hanya 66%, KNN yang diimplementasikan dalam penelitian ini dianggap efektif karena mampu mengidentifikasi variabel penting yang mempengaruhi kelulusan dan memberikan wawasan berharga bagi pengambilan keputusan dalam institusi pendidikan.

## DAFTAR PUSTAKA

- [1] F. Asrory, J. D. Irawan, and A. Wahid, "Aplikasi Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode KNN," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 4, no. 2, pp. 105–112, Oct. 2020, doi: 10.36040/jati.v4i2.2725.
- [2] Z. Amri, K. Kusriani, and K. Kusnawi, "Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM," *Edumatic: Jurnal Pendidikan Informatika*, vol. 7, no. 2,

- pp. 187–196, Dec. 2023, doi: 10.29408/edumatic.v7i2.18620.
- [3] S. Widaningsih, “Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naive Bayes, KNN dan SVM,” *Jurnal Tekno Insentif*, vol. 13, no. 1, pp. 16–25, Apr. 2019, doi: 10.36787/jti.v13i1.78.
- [4] J. Astri, J. Karman, and N. K. Daulay, “Prediksi Kelulusan Mahasiswa Menggunakan Metode K-Nearest Neighbor (KNN) pada Fakultas Ilmu Teknik, Univeritas Bina Insan,” vol. 8, pp. 169–173, 2023, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- [5] O. Sativa, D. Putri, A. Yhurinda, P. Putri, I. T. Adhi, and T. Surabaya, “SNESTIK Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika Prediksi Kelulusan Mahasiswa dengan Metode K-Nearest Neighbor pada Jurusan Sistem Informasi Institut XYZ,” 2023, doi: 10.31284/p.snestik.2023.4388.
- [6] D. Prasetyawan, R. Gatra, K. Kunci, P. Akademik, and K. Pearson, “Algoritma K-Nearest Neighbor untuk Memprediksi Prestasi Mahasiswa Berdasarkan Latar Belakang Pendidikan dan Ekonomi,” 2022.
- [7] A. Ulul Azmi Wafiqi, T. Tundo, B. Arvian James, A. Huga Ramadhan, and A. Nizar, “Prediksi Tingkat Stres Pada Mahasiswa UNUGHA Cilacap Menggunakan Algoritma K-Nearest Neighbor,” *Jurnal TEKNO KOMPAK*, vol. 18, no. 2, 2024.
- [8] R. Ridwan, H. Lubis, and P. Kustanto, “Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa,” *Jurnal Media Informatika Budidarma*, vol. 4, no. 2, p. 286, Apr. 2020, doi: 10.30865/mib.v4i2.2035.
- [9] O. Bangun, H. Mawengkang, and S. Efendi, “Metode Algoritma Support Vector Machine (SVM) Linier Dalam Memprediksi Kelulusan Mahasiswa,” *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, p. 2006, Oct. 2022, doi: 10.30865/mib.v6i4.4572.
- [10] M. D. Yalidhan, “Implementasi Algoritma Backpropagation Untuk Memprediksi Kelulusan Mahasiswa,” *Klik-Kumpulan Jurnal Ilmu Komputer*, vol. 5, no. 2, p. 169, Sep. 2018, doi: 10.20527/klik.v5i2.152.
- [11] A. Pratama, R. Cahya Wihandika, and D. E. Ratnawati, “Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa,” 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>

- [12] R. Puspita, S. Putri, I. Waspada, and U. D. Semarang, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," 2018.
- [13] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic Review," *Faktor Exacta*, vol. 13, no. 1, p. 35, Jun. 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [14] L. Abd, R. Hakim, A. A. Rizal, and D. Ratnasari, "Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN)," 2019.
- [15] A. Yandi Saputra and Y. Primadasa, "Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour Implementation of Classification Method to Predict Student Graduation Using K-Nearest Neighbor Algorithm," 2018.
- [16] O. Caelen, "A Bayesian interpretation of the confusion matrix," *Ann Math Artif Intell*, vol. 81, no. 3–4, pp. 429–450, Dec. 2017, doi: 10.1007/s10472-017-9564-8.