

Penggunaan Algoritma K-Nearest Neighbors (K-NN) untuk Mengklasifikasi Stres Akademik Mahasiswa Departemen Matematika Universitas Negeri Padang

Muhamad Ihsan, Meira Parma Dewi

Departemen Matematika, Universitas Negeri Padang

Article Info

Article history:

Received Agustus 11, 2025

Revised Agustus 27, 2025

Accepted September 10, 2025

Keywords:

Academic Stress

Classification

K-NN

Prediction

Data Mining

Kata Kunci:

Stres Akademik

Klasifikasi

K-NN

Prediksi

Data Mining

ABSTRACT

This study aims to classify the academic stress levels of students in the Mathematics Department of Padang State University using the K-Nearest Neighbors (K-NN) algorithm. The research data were obtained from 100 students selected using the Stratified Proportional Random Sampling method. The variables used include study program, year, gender, GPA, number of semester credits units, and Student Stress Inventory (SSI) score. The analysis process includes data pre-processing, selecting the optimal K parameter using grid search, and model evaluation using a confusion matrix. The results showed that K-NN with $K = 7$ produced an accuracy of 95%, precision of 95%, recall of 100%, and an F1-score of 97.4%, with the majority of students being in the mild to moderate stress category. These findings indicate that K-NN is effective for classifying students' academic stress levels and can be a basis for the university in designing targeted stress management programs.

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan tingkat stres akademik mahasiswa Departemen Matematika Universitas Negeri Padang menggunakan algoritma K-Nearest Neighbors (K-NN). Data penelitian diperoleh dari 100 mahasiswa yang dipilih dengan metode *Stratified Proportional Random Sampling*. Variabel yang digunakan meliputi program studi, angkatan, jenis kelamin, IPK, jumlah SKS, dan skor *Student Stress Inventory* (SSI). Proses analisis meliputi praprosesan data, pemilihan parameter K optimal menggunakan grid search, dan evaluasi model menggunakan *confusion matrix*. Hasil penelitian menunjukkan bahwa K-NN dengan $K=7$ menghasilkan akurasi sebesar 95%, presisi 95%, recall 100%, dan F1-score 97,4%, dengan mayoritas mahasiswa berada pada kategori stres ringan hingga sedang. Temuan ini menunjukkan bahwa K-NN efektif digunakan untuk mengklasifikasikan tingkat stres akademik mahasiswa, serta dapat menjadi dasar bagi pihak kampus dalam merancang program manajemen stres yang tepat sasaran.

This is an open access article under the [CC BY-SA](#) license.



Penulis Korespondensi:

Muhamad Ihsan

Departemen Matematika, Universitas Negeri Padang,

Email: muhammadihsan12378@gmail.com

1. PENDAHULUAN

Mahasiswa merupakan individu yang sedang mengikuti tahapan atau proses pendidikan di salah satu perguruan tinggi yang dimasukinya semakin tinggi tingkatan ataupun semester yang dijalani seorang mahasiswa diharapkan semakin mampu memahami pelajaran serta mencari solusi terbaik disetiap permasalahan yang ditemui [1]. Pada masa ini tidak jarang mahasiswa mengalami tekanan baik dari dirinya sendiri (internal) ataupun dari lingkungannya (eksternal). Salah satu yang paling dirasakan adalah stres yang terkait dengan aktivitas akademik. Mahasiswa adalah individu yang sangat mudah dan rentan mengalami stres. Stres merupakan respon seseorang terhadap stressor (kejadian, faktor, atau stimulus yang memicu datangnya stres terhadap seseorang). Stres muncul akibat tekanan yang datang dalam bentuk tuntutan yang harus dan segera dikerjakan oleh seorang mahasiswa [2]. Stres akademik merupakan suatu kondisi psikologis yang kerap dialami oleh mahasiswa selama menjalani studi. Stres ini dapat disebabkan oleh tekanan internal maupun eksternal yang bersumber dari tuntutan akademik, beban tugas, maupun lingkungan sosial di kampus [3]. Masalah stres akademik sering kali tidak disadari oleh mahasiswa itu sendiri sehingga diperlukan suatu metode untuk mendeteksi tingkat stres yang dialami mahasiswa.

Salah satu metode yang dapat digunakan untuk mendeteksi tingkat stres akademik adalah algoritma *K-Nearest Neighbors* (K-NN). Algoritma ini bekerja dengan mencari data *training* yang memiliki jarak terdekat dengan data uji berdasarkan nilai atribut tertentu [4]. Algoritma K-NN dipilih dikarenakan algoritma ini memiliki beberapa keunggulan yang relevan seperti merupakan algoritma sederhana yang mudah dimengerti oleh banyak orang, serta hasil prediksi dari algoritma K-NN ini juga cukup akurat meski hanya menggunakan parameter yang minimal. Karena algoritma K-NN dikenal fleksibel dalam mengolah data baik data numerik (data yang berbentuk angka) maupun data kategorik (data yang berbentuk non-numerik) [5]. Dalam penelitian ini, data yang digunakan adalah data mahasiswa Departemen Matematika Universitas Negeri Padang tahun masuk 2021 hingga 2024. Atribut yang digunakan meliputi program studi, tahun masuk, jenis kelamin, indeks prestasi kumulatif (IPK), jumlah satuan kredit semester (SKS), dan hasil kuesioner *Student Stress Inventory* (SSI) yang terdiri dari empat subskala: fisik, interpersonal, akademik, dan lingkungan [6].

Tujuan dari penelitian ini adalah untuk mengetahui hasil klasifikasi untuk memprediksi tingkat stres akademik mahasiswa Departemen Matematika Universitas Negeri Padang dengan menggunakan algoritma K-NN, serta mengevaluasi hasil akurasi klasifikasinya. Penelitian ini diharapkan mampu memberikan hasil klasifikasi tingkat stres akademik mahasiswa Departemen Matematika yang akurat. Berdasarkan latar belakang masalah yang telah dibuat diperoleh perumusan masalah adalah “Bagaimana Penggunaan Algoritma *K-Nearest Neighbors* (K-NN) untuk Mengklasifikasi Tingkat Stres Akademik Mahasiswa Departemen Matematika Universitas Negeri Padang ?”

2. METODE

Jenis penelitian ini adalah penelitian terapan dengan pendekatan kuantitatif, yang bertujuan untuk mengklasifikasi tingkat stres akademik mahasiswa tahun masuk 2021 hingga 2024 di Departemen Matematika Universitas Negeri Padang yang meliputi program studi Matematika, Program Studi Pendidikan Matematika, dan Aktuaria. Penelitian terapan yaitu penelitian yang bertujuan untuk memecahkan atau memberikan solusi atas suatu permasalahan. Penelitian ini diawali dengan analisis teori kemudian diikuti dengan pengambilan data dan analisis data.

2.1 Populasi dan Sampel

Populasi adalah semua individu atau objek yang memiliki karakteristik umum yang sama. Populasi dalam penelitian ini adalah mahasiswa tahun masuk 2021 sampai tahun masuk 2024 Departemen Matematika Universitas Negeri Padang. Sampel adalah bagian dari elemen populasi, sampel diambil dengan melalui proses memilih sebagian dari populasi untuk mewakili keseluruhan populasi. Teknik pengambilan sampel pada penelitian ini

menggunakan teknik *Stratified Proportional Random Sampling*. Metode ini dilakukan dengan cara membagi populasi ke dalam beberapa strata berdasarkan program studi dan angkatan. Kemudian, dari setiap strata tersebut diambil sampel secara acak dan proporsional sesuai dengan ukuran populasi masing-masing strata. Dengan kriteria responden dalam penelitian ini yaitu: Mahasiswa tahun masuk 2021, 2022, 2023, dan 2024 di Departemen Matematika Universitas Negeri Padang, yang meliputi program studi Matematika, Program Studi Pendidikan Matematika, dan Aktuaria. Dalam penelitian ini ditetapkan jumlah sampel sebesar 100 mahasiswa.

2.2 Variabel

Variabel penelitian pada dasarnya adalah suatu hal yang berbentuk apa saja yang ditetapkan oleh peneliti untuk mempelajari sehingga dapat memperoleh informasi tentang hal tersebut. Variabel penelitian terdiri dari :

- a. Variabel dependen (terikat)

Variabel dependen yang digunakan untuk memprediksi tingkat stres mahasiswa semester akhir di Departemen Matematika Universitas Negeri Padang adalah tingkat stres mahasiswa menggunakan Skor SSI (Student Stress Inventory) untuk menentukan kelas.

- b. Variabel independen (bebas)

Variabel independen yang digunakan dalam penelitian ini antara lain :

1. Program studi (X_1)
2. Tahun masuk (X_2)
3. Jenis kelamin (X_3)
4. Indeks Prestasi Kumulatif / IPK (X_4)
5. Jumlah SKS (X_5)
6. Total skor SSI (X_6)

2.3 Instrumen Penelitian

Instrumen yang digunakan adalah angket berupa kuesioner yang berisi masalah yang akan diteliti. Kuesioner diadaptasi dari SSI (*Student Stress Inventory*) yang sudah tervalidasi oleh *Fakulty of Education and Human Development Sultan Idris Education University*. Kuesioner yang digunakan dalam penelitian ini terdiri dari 40 pertanyaan, dimana terbagi menjadi 4 subskala yaitu fisik, hubungan interpersonal, akademik, dan lingkungan. Yang masing-masing subskala terdiri dari 10 item pertanyaan. Sesuai pada kuesioner yang digunakan tingkatan stres dibagi menjadi 3 kelas yaitu Stres : Ringan, Sedang, Berat.

2.4 Teknik Analisis Data

Adapun tahapan analisis data adalah sebagai berikut:

1. Pengumpulan dataset.
2. *Preprocessing* data, yaitu proses *cleaning* atau pembersihan data pendukuk yang tidak dipakai kemudian menentukan data yang digunakan sebagai atribut untuk melakukan perhitungan sehingga menghasilkan dataset baru [7].
3. Pembagian data menjadi data *taining* dan data *testing*.
4. Melakukan normalisasi data dengan melakukan standarisasi terhadap data numerik.
5. Menentukan nilai K optimal.
6. Penerapan K-NN dengan menghitung jarak menggunakan *Euclidean Distance*, kemudian mengurutkannya berdasarkan jarak dari yang terkecil hingga jarak terbesar.
7. Menentukan tetangga terdekat berdasarkan nilai K.
8. Melakukan evaluasi pada data *testing* menggunakan *Confusion Matriks* untuk melihat akurasi dalam mengklasifikasi stres akademik mahasiswa [8].

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data

Data yang digunakan dalam penelitian ini diperoleh dari data primer dengan menyebarkan kuesioner kepada mahasiswa aktif Departemen Matematika Universitas Negeri Padang, tahun masuk 2021, 2022, 2023, dan 2024. Data dikumpulkan berdasarkan skala *Likert* dari kuesioner *Student Stress Inventory* (SSI). SSI dibagi menjadi empat subskala: stres fisik, stres interpersonal, stres akademis, dan stres lingkungan. Sampel terdiri dari 100 mahasiswa. Data dikumpulkan mencakup informasi mengenai mahasiswa seperti NIM, Prodi, Jenis Kelamin, Jumlah SKS, IPK, serta skor SSI. Berikut merupakan data yang diperoleh:

Tabel 1. Contoh Dataset Mahasiswa

No.	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	Y
1	Matematika	2024	Perempuan	3,25	19	75	SR
2	Matematika	2021	Perempuan	3,2	146	106	SS
3	Pendidikan Matematika	2024	Perempuan	3,5	21	84	SS
4	Pendidikan Matematika	2022	Perempuan	3,1	104	108	SS
5	Aktuaria	2024	Perempuan	2,76	20	96	SS
6	Pendidikan Matematika	2022	Laki-laki	3,33	102	70	SR
7	Matematika	2021	Laki-laki	3,69	138	72	SR
8	Pendidikan Matematika	2023	Laki-laki	3,66	66	99	SS
9	Pendidikan Matematika	2021	Laki-laki	3,8	146	80	SR
10	Matematika	2022	Perempuan	3,11	98	88	SS
:	:	:	:	:	:	:	:
90	Aktuaria	2024	Laki-laki	3,62	20	45	SR
91	Pendidikan Matematika	2021	Perempuan	3,48	144	72	SR
92	Pendidikan Matematika	2024	Perempuan	3,86	21	89	SS
93	Pendidikan Matematika	2023	Perempuan	3,44	85	71	SR
94	Pendidikan Matematika	2023	Laki-laki	3,4	85	77	SR
95	Matematika	2024	Perempuan	2,9	19	123	SB
96	Pendidikan Matematika	2022	Perempuan	3,45	109	107	SS
97	Matematika	2024	Perempuan	2,81	19	89	SS
98	Matematika	2024	Perempuan	3,12	19	75	SR
99	Matematika	2021	Perempuan	3,21	141	100	SS
100	Pendidikan Matematika	2021	Perempuan	3,28	141	83	SS

Data pada Tabel 1 merupakan contoh *dataset* yang akan digunakan dalam proses penerapan algoritma K-NN yang akan menampilkan NIM, Prodi (X₁), Tahun Masuk (X₂), Jenis Kelamin (X₃), dan fitur numerik seperti: IPK (X₄), Jumlah SKS (X₅), dan Total Skor SSI (X₆). Hasil pada kolom X₆ didapatkan dengan menjumlahkan nilai seluruh item pertanyaan pada kuesioner yang berskala *likert* 4 dari masing-masing responden.

3.2 Pengolahan Data

Dari data yang telah diperoleh akan dilakukan pengolahan data dengan tahapan pra-pemrosesan data (*preprocessing*), yang meliputi: data *cleaning*, menghapus data yang hilang (*missing values*) dan transformasi data. Proses ini dilakukan untuk mengubah suatu nilai atribut menjadi format yang sesuai [9].

Tabel 2. Rentang Stres

Total Skor	Tingkat Stres
40-80	Stres Ringan
81-121	Stres Sedang
122-160	Stres Berat

Data numerik akan dinormalisasi agar rentangnya seragam dan efektif untuk perhitungan jarak menggunakan jarak *Euclidean distance* dalam algoritma K-NN. Algoritma K-NN memanfaatkan data sampel yang sudah dilabeli, terdiri atas sejumlah variabel prediktor dan kelas target pada data *training*. Sementara itu pada data *testing*, K-NN menghitung jarak antara setiap data *testing* dengan seluruh data *training* berdasarkan nilai atribut-atributnya menggunakan rumus *Euclidean Distance*, untuk menentukan kelas berdasarkan mayoritas tertinggi terdekat [10]. *Dataset* yang digunakan berjumlah 100 data, dan data tersebut terdiri dari 80 data *training* dan 20 data *testing*. Berikut merupakan sebagian sampel data *training*.

Tabel 3. Sampel Data Training

No.	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	Y
1	1		3	1	3,25	19	75 SR
2	1		0	1	3,2	146	106 SS
3	2		3	1	3,5	21	84 SS
4	2		1	1	3,1	104	108 SS
5	0		3	1	2,76	20	96 SS
6	2		1	0	3,33	102	70 SR
7	1		0	0	3,69	138	72 SR
8	2		2	0	3,66	66	99 SS
9	2		0	0	3,8	146	80 SR
10	1		1	1	3,11	98	88 SS
⋮	⋮		⋮	⋮	⋮	⋮	⋮
76	2		3	1	3,17	20	77 SR
77	2		1	0	3,13	102	79 SR
78	2		0	1	3,65	140	91 SS
79	2		1	0	3,06	101	74 SR
80	2		2	1	3,44	86	79 SR

Sesuai kuesioner yang digunakan dari SSI (*Student Stress Inventory*) yang sudah tervalidasi oleh *Fakulty of Education and Human Development Sultan Idris Education University*, tingkatan stres dibagi menjadi 3 kelas yaitu, Stres Berat, Stres Sedang, dan Stres Ringan dimana rentang tingkatan stresnya dapat dilihat pada Tabel 2.

Tabel 4. Sampel Data Testing

No.	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	Y
1	2	2	1	3,3	92	94	SS
2	1	3	1	3,2	19	82	SS
3	1	3	1	3,36	19	95	SS
4	1	2	1	3	83	94	SS
5	2	0	1	3,4	145	79	SR
6	2	0	1	3,61	142	79	SR
7	1	0	1	3,57	141	64	SR
8	1	2	1	3,16	63	83	SS
9	2	0	1	3,36	140	76	SR
10	0	3	0	3,62	20	45	SR
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	2	1	1	3,45	109	107	SS
17	1	3	1	2,81	19	89	SS
18	1	3	1	3,12	19	75	SR
19	1	0	1	3,21	141	100	SS
20	2	0	1	3,28	141	83	SS

Rentang stres tersebut berfungsi sebagai criteria mapping, sebagai acuan mengkonversi hasil perhitungan skor stress menjadi kategori klasifikasi.

Tabel 3 menampilkan sampel data *training* yang akan digunakan dalam penelitian ini. Selanjutnya di tampilkan juga data *testing* pada panelitian ini pada Tabel 4. Dari data *training* dan data *testing* yang sudah ditampilkan tersebut, langkah selanjutnya adalah melakukan normalisasi data dengan melakukan standarisasi terhadap data numerik (X_4, X_5, X_6). berikut merupakan tabel hasil standarisasi data.

Tabel 5. Sampel Data Training Hasil Normalisasi

No.	X_1	X_2	X_3	X_4	X_5	X_6	Y
1	1	3	1	-0,256985067	0,138912119	0,715830885	SR
2	1	0	1	-0,658524235	-1,454892943	-0,128392272	SS
3	2	3	1	-0,016061567	-1,454892943	0,786182814	SS
4	2	1	1	-1,46160257	-0,057584395	0,715830885	SS
5	0	3	1	0,1445541	1,29605826	-0,339448061	SS
6	2	1	0	0,987786352	1,230559422	-0,339448061	SR
7	1	0	0	0,827170685	1,208726476	-1,394727006	SR
8	2	2	0	-0,819139902	-0,494243316	-0,058040342	SS
9	2	0	0	-0,016061567	1,18689353	-0,55050385	SR
10	1	1	1	1,027940269	-1,433059997	-2,731413671	SS
:	:	:	:	:	:	:	:
16	2	1	1	0,345323684	0,510072202	1,630405971	SS
17	1	3	1	-2,224526989	-1,454892943	0,364071236	SS
18	1	3	1	-0,979755569	-1,454892943	-0,62085578	SR
19	1	0	1	-0,618370318	1,208726476	1,137942463	SS
20	2	0	1	-0,337292901	1,208726476	-0,058040342	SS

Tabel 5 menampilkan sampel dari data *training* yang sudah dilakukan normalisasi, selanjutnya juga akan ditampilkan sampel data *testing* yang sudah dilakukan standarisasi data. Sedangkan untuk sampel dari data *testing* akan ditampilkan pada Tabel 6.

Tabel 6. Sample Data Testing Hasil Normalisasi

No.	X_1	X_2	X_3	X_4	X_5	X_6	Y
1	2	2	1	-0,256985067	0,138912119	0,715830885	SS
2	1	3	1	-0,658524235	-1,454892943	-0,128392272	SS
3	1	3	1	-0,016061567	-1,454892943	0,786182814	SS
4	1	2	1	-1,46160257	-0,057584395	0,715830885	SS
5	2	0	1	0,1445541	1,29605826	-0,339448061	SR
6	2	0	1	0,987786352	1,230559422	-0,339448061	SR
7	1	0	1	0,827170685	1,208726476	-1,394727006	SR
8	1	2	1	-0,819139902	-0,494243316	-0,058040342	SS
9	2	0	1	-0,016061567	1,18689353	-0,55050385	SR
10	0	3	0	1,027940269	-1,433059997	-2,731413671	SR
:	:	:	:	:	:	:	:
16	2	1	1	0,345323684	0,510072202	1,630405971	SS
17	1	3	1	-2,224526989	-1,454892943	0,364071236	SS
18	1	3	1	-0,979755569	-1,454892943	-0,62085578	SR
19	1	0	1	-0,618370318	1,208726476	1,137942463	SS
20	2	0	1	-0,337292901	1,208726476	-0,058040342	SS

3.3 Menentukan K Optimal

Dalam algoritma K-NN, pemilihan nilai parameter K sangat mempengaruhi kinerja model, jika K terlalu kecil bisa menyebabkan *overfitting* (sensitif terhadap noise) dan jika nilai K juga terlalu besar bisa menyebabkan *underfitting* [11]. Karena itu perlu dilakukannya pengujian K untuk memperoleh akurasi yang tinggi dengan error yang rendah. Berikut ini merupakan pengujian akurasi menggunakan metode *Grid Search*, merupakan sebuah grafik yang menunjukkan hubungan antara nilai K dengan akurasi model K-NN [12].



Gambar 1. Grafik Akurasi

Selanjutnya dalam proses klasifikasi nilai parameter K digunakan untuk menentukan jumlah tetangga terdekat. Titik-titik biru (plot) pada gambar menunjukkan seberapa besar nilai akurasi. Pada penelitian ini nilai parameter K yang akan digunakan adalah **K = 7**, karena berdasarkan hasil dari grafik akurasi menunjukkan hasil akurasi tertinggi dengan nilai akurasi sebesar **95%**.

3.4 Penerapan Algoritma K-NN

Penerapan algoritma K-NN diawali dengan proses klasifikasi dilakukan dengan menghitung jarak antara data *testing* dan data *training* menggunakan rumus pengukuran jarak *Euclidean Distance*. Selanjutnya hasil perhitungan jarak dapat dilihat pada tabel berikut:

Tabel 7. Sampel Hasil Perhitungan Jarak *Euclid* Data *Testing* Terhadap Data *Training*

Data test 1			Data test 2			...	Data test 20		
Data Training	Jarak Euclid	Urutan	Data Training	Jarak Euclid	Urutan	...	Data Training	Jarak Euclid	Urutan
64	0.507528	1	59	0.467426	1	...	60	0.752395	1
14	1.010464	2	1	0.531816	2	...	39	1.017255	2
66	1.052404	3	41	0.874596	3	...	29	1.026027	3
30	1.054936	4	76	1.067109	4	...	43	1.038981	4
24	1.099859	5	32	1.102114	5	...	65	1.190512	5
80	1.202827	6	31	1.150516	6	...	57	1.197844	6
25	1.268201	7	75	1.156599	7	...	15	1.232422	7
15	1.300286	8	74	1.250045	8	...	35	1.245792	8
35	1.316896	9	69	1.265727	9	...	13	1.258611	9
57	1.461491	10	45	1.39288	10	...	18	1.330762	10
1	1	1	1	1	1	...	1	1	1
9	3.375043	74	78	4.543668	74	...	44	4.279026	74
7	3.44333	75	33	4.54669	75	...	41	4.301345	75
11	3.473539	76	7	4.595257	76	...	28	4.342246	76
5	3.491772	77	19	4.720347	77	...	73	4.501362	77
19	3.529973	78	9	4.951014	78	...	11	4.695492	78
53	3.607912	79	26	5.228649	79	...	50	4.839526	79
50	3.776785	80	53	5.361017	80	...	5	5.017494	80

Tabel 7 menampilkan beberapa data hasil perhitungan jarak *euclid* sampel data *testing* terhadap data *training*. Setelah melakukan perhitungan jarak, langkah berikutnya yaitu memberikan label kelas prediksi data *testing* berdasarkan nilai K dan jarak terdekat.

Tabel 8. Sampel Hasil Prediksi Data Testing Terhadap Data Training

Data test 1				Data test 2				...				Data test 20					
Data Training	Jarak Euclid	Urutan	Kelas	Data Training	Jarak Euclid	Urutan	Kelas	...	Data Training	Jarak Euclid	Urutan	Kelas	...	Data Training	Jarak Euclid	Urutan	Kelas
64	0,507528	1	SS	59	0,467426	1	SS	...	60	0,752395	1	SR
14	1,010464	2	SS	1	0,531816	2	SR	...	39	1,017255	2	SS
66	1,052404	3	SS	41	0,874596	3	SR	...	29	1,026027	3	SS
30	1,054936	4	SS	76	1,067109	4	SR	...	43	1,038981	4	SS
24	1,099859	5	SS	32	1,102114	5	SS	...	65	1,190512	5	SR
80	1,202827	6	SR	31	1,150516	6	SS	...	57	1,197844	6	SS
25	1,268201	7	SR	75	1,156599	7	SS	...	15	1,232422	7	SS
15	1,300286	8	SS	74	1,250045	8	SR	...	35	1,245792	8	SS
35	1,316896	9	SS	69	1,265727	9	SS	...	13	1,258611	9	SS
57	1,461491	10	SS	45	1,39288	10	SR	...	18	1,330762	10	SR
...
9	3,375043	74	SR	78	4,543668	74	SS	...	44	4,279026	74	SS
7	3,44333	75	SR	33	4,54669	75	SR	...	41	4,301345	75	SR
11	3,473539	76	SR	7	4,595257	76	SR	...	28	4,342246	76	SR
5	3,491772	77	SS	19	4,720347	77	SR	...	73	4,501362	77	SR
19	3,529973	78	SR	9	4,951014	78	SR	...	11	4,695492	78	SR
53	3,607912	79	SB	26	5,228649	79	SB	...	50	4,839526	79	SR
50	3,776785	80	SR	53	5,361017	80	SB	...	5	5,017494	80	SS
	Prediksi	SS		Prediksi	SS			...	Prediksi	SS							

Data aktual dari hasil dari prediksi data *testing* keseluruhan beserta perbandingan kelas data *testing* asli dapat dilihat pada tabel dibawah ini.

Tabel 9. Hasil Prediksi Seluruh Data Testing

No.	Data Testing	Data Aktual	Data Prediksi	Keterangan
1	1	SS	SS	Sesuai
2	2	SS	SS	Sesuai
3	3	SS	SS	Sesuai
4	4	SS	SS	Sesuai
5	5	SR	SR	Sesuai
6	6	SR	SR	Sesuai
7	7	SR	SR	Sesuai
8	8	SS	SS	Sesuai
9	9	SR	SR	Sesuai
10	10	SR	SR	Sesuai
11	11	SR	SR	Sesuai
12	12	SS	SS	Sesuai
13	13	SR	SR	Sesuai
14	14	SR	SR	Sesuai
15	15	SB	SS	Tidak Sesuai
16	16	SS	SS	Sesuai
17	17	SS	SS	Sesuai
18	18	SR	SR	Sesuai
19	19	SS	SS	Sesuai
20	20	SS	SS	Sesuai

Data aktual (label dari kelas sesungguhnya) didapat dari Tabel 7 sebelumnya telah diketahui dan data prediksi didapat dari hasil kelas yang diprediksi oleh model. Hasil analisis dari total data *testing* dengan total 20 data dapat diprediksi dengan benar (sesuai) sebanyak 19 data dan diprediksi salah (tidak sesuai) sebanyak 1 data, yaitu pada data *testing* ke-15 dengan data aktual SB yang diprediksi SS.

4 Evaluasi Model K-NN

Untuk mengetahui kinerja dari algoritma K-NN dalam melakukan klasifikasi terhadap suatu kelas/label yang sudah ditentukan, akan dilakukan pengujian pada hasil akurasi [13]. Hasil dari pengujian akurasi akan dimunculkan dalam bentuk Confusion Matrix 3×3 pada tabel berikut:

Tabel 10. Confusion Matrix 3×3 hasil klasifikasi K-NN

Confusion Matrix		Prediksi		
		SR	SS	SB
Aktual	SR	9	0	0
	SS	0	10	0
	SB	0	1	0

Pada tabel diketahui bahwa *Confusion Matrix* menunjukkan performa dari model K-NN dalam mengklasifikasi *dataset* ke dalam tiga kelas (SR, SS, SB). Model berhasil mengklasifikasikan sebanyak 9 data SR dengan benar, dan 10 data SS yang juga diprediksi benar. Namun, model melakukan satu kesalahan prediksi dimana terdapat 1 data SB yang diklasifikasi sebagai SS. Kesimpulannya dari total 20 data *testing*, 19 data diklasifikasikan dengan benar, dan 1 data diklasifikasikan dengan salah.

Tabel 11. Confusion Matrix Biner 2×2 Hasil klasifikasi K-NN

		Nilai Aktual	
		Positif	Negatif
Nilai Prediksi	Positif	19	1
	Negatif	0	0

Selanjutnya pada Tabel 11 menunjukkan *Confusion Matrix* versi biner dari klasifikasi multi-kelas pada Tabel 10 sebelumnya. Pada *Confusion Matrix* ini, hasil diklasifikasikan sebagai positif (klasifikasi benar), dan negatif (klasifikasi salah) [14], hasilnya sebagai berikut:

- *True Positive (TP) = 19*
- *False Positive (FP) = 1*
- *False Negative (FN) 0*
- *True Negative (TN) = 0*

Berikut akan ditampilkan hasil perhitungannya:

1. Akurasi (*Accuracy*):

$$\begin{aligned} \text{akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\ &= \frac{19 + 0}{19 + 0 + 1 + 0} = \frac{19}{20} = 0,95 \times 100\% = 95\% \end{aligned}$$

Model mampu mengklasifikasikan 95% dari total data *testing*, Artinya, 19 data diklasifikasi benar, namun ada 1 data salah diklasifikasikan.

2. Presisi (*Precision*):

$$\begin{aligned} \text{presisi} &= \frac{TP}{TP + FP} \times 100\% \\ &= \frac{19}{19 + 1} = \frac{19}{20} = 0,95 \times 100\% = 95\% \end{aligned}$$

Artinya 95% hasil prediksi dianggap benar berdasarkan data aktual, dan dengan 5% tingkat kesalahan hasil prediksi yang dilakukan oleh model. Hal ini menunjukkan tingkat kepercayaan terhadap hasil prediksi cukup tinggi.

3. Recall (*Sensitivity*):

$$\begin{aligned} \text{recall} &= \frac{TP}{TP + FN} \times 100\% \\ &= \frac{19}{19 + 0} = \frac{19}{19} = 1 \times 100\% = 100\% \end{aligned}$$

Jumlah data aktual yang diprediksi dengan benar menggunakan model.

4. F1-Score:

$$\begin{aligned} F1 &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \times 100\% \\ &= 2 \times \frac{0,95 \times 1}{0,95 + 1} = \frac{1,9}{1,95} \approx 0,974 \times 100\% = 97,4\% \end{aligned}$$

F1-Score menyeimbangkan nilai presisi dan recall [15]. Berdasarkan hasil yang didapatkan di atas diperoleh diketahui nilai F1-Score sebesar 97,4% menunjukkan bahwa model konsisten dalam mengenali seluruh data positif secara lengkap

5. Error Rate:

$$\begin{aligned} \text{error} &= \frac{FP + FN}{TP + TN + FP + FN} \times 100\% \\ &= \frac{1 + 0}{20} = 0,05 \times 100\% = 5\% \end{aligned}$$

Didapatkan hasil *error rate* sebesar 5%, yang berarti hasil prediksi dengan model hanya memiliki kesalahan sebesar 5% dalam mengklasifikasikan data.

4. KESIMPULAN

Setelah menggunakan Algoritma K-NN untuk mengklasifikasikan tingkat stres akademik dalam lingkup Departemen Matematika di Universitas Negeri Padang, didapatkan hasil penelitian serta pembahasan yang dilakukan dalam penelitian ini berhasil mencapai dua tujuan utama penelitian ini seperti sebagai berikut :

1. Penelitian ini menghasilkan klasifikasi tingkat stres akademik mahasiswa Departemen Matematika Universitas Negeri Padang ke dalam tiga kategori yaitu stres ringan, stres sedang, dan stres berat. dengan menggunakan algoritma K-NN. Klasifikasi yang dihasilkan menunjukkan bahwa mayoritas mahasiswa berada pada kategori stres ringan dan sedang, sementara sebagian kecil berada pada kategori stres berat.
2. Hasil evaluasi menggunakan *confusion matrix* dan implementasi model di *Google Colab*, algoritma K-NN menunjukkan performa yang sangat baik dengan tingkat akurasi sebesar 95%. Hal ini menunjukkan bahwa algoritma K-NN sangat efektif dan akurat dalam mengklasifikasi tingkat stres akademik mahasiswa, serta dapat dijadikan alat bantu dalam pengambilan keputusan di lingkungan pendidikan, khususnya dalam upaya memantau dan menangani kesehatan mental mahasiswa.

REFERENSI

- [1] I. K. Siregar and S. R. Putri, "Hubungan Self-Efficacy dan Stres Akademik Mahasiswa," *Cons. Berk. Kaji. Konseling dan Ilmu Keagamaan*, vol. 6, no. 2, p. 91, 2020, doi: 10.37064/consilium.v6i2.6386.
- [2] H. C. Merry; Mamahit, "Stres Akademik Mahasiswa Aktif Angkatan 2018 dan 2019 Universitas Swasta di DKI Jakarta," *J. Konseling Indones.*, vol. 6, no. 1, pp. 6–13, 2020, [Online]. Available: <http://ejournal.unikama.ac.id/index.php/JKI>
- [3] N. Mawakhira Yusuf and J. Ma'wa Yusuf, "Faktor-Faktor Yang Mempengaruhi Stres Akademik," *Psyche 165 J.*, vol. 13, no. 02, pp. 235–239, 2020, [Online]. Available: <https://jpsy165.org/ojs/index.php/jpsy165/article/view/84>
- [4] K. Kartarina, N. K. Sriwinarti, and N. Iuh P. Juniarti, "Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam

- Memprediksi Kelulusan Mahasiswa," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 2, pp. 107–113, 2021, doi: 10.35746/jtim.v3i2.159.
- [5] M. N. Maskuri, K. Sukerti, and R. M. Herdian Bhakti, "Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke Stroke Disease Predict Using KNN Algorithm," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 1, pp. 130–140, 2022.
- [6] M. A. S. Mohamed Arip, D. N. Kamaruzaman, A. Roslan, A. Ahmad, and M. R. Abd Rahman, "Development, validity and reliability of student stress inventory (SSI)," *Soc. Sci.*, vol. 10, no. 7, pp. 1631–1638, 2015, doi: 10.3923/sscence.2015.1631.1638.
- [7] I. dan A. Mutiara, "Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4," *Klik - Kumpul. J. Ilmu Komput.*, vol. 2, no. 2, pp. 159–173, 2015.
- [8] Mustakim and G. Oktaviani F, "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa," vol. 13, no. 2, pp. 195–202, 2016.
- [9] B. G. Sudarsono and S. P. Lestari, "Diagnosa Tingkat Depresi Mahasiswa Akhir Terhadap Penelitian Ilmiah Menggunakan Algoritma K-Nearest Neighbor," *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 1094–1099, 2020, doi: 10.30865/mib.v4i4.2448.
- [10] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 6, no. 2, pp. 118–127, 2021, doi: 10.31294/ijcit.v6i2.10438.
- [11] D. Widjadha, R. B. Hastuti, I. Kharisudin, and F. Fauzi, "Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah," *Prism. Pros. Semin. Nas. Mat.*, vol. 4, no. 2, pp. 584–594, 2021.
- [12] Sukamto, Hadiyanto, and Kurnianingsih, "KNN Optimization Using Grid Search Algorithm for Preeclampsia Imbalance Class," *E3S Web Conf.*, vol. 448, 2023, doi: 10.1051/e3sconf/202344802057.
- [13] S. Sathyaranayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African J. Biomed. Res.*, no. November, pp. 4023–4031, 2024, doi: 10.53555/ajbr.v27i4s.4345.
- [14] J. Supriyanto, D. Alita, and A. R. Isnain, "Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, pp. 74–80, 2023, doi: 10.33365/jatika.v4i1.2468.
- [15] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.