

SISTEM KLASIFIKASI UNTUK MENENTUKAN TINGKAT STRESS MAHASISWA SECARA UMUM MENGGUNAKAN METODE K-NEAREST NEIGHBORS

Sopwatun Anisa^{1*}, Agus Komarudin², Edvin Ramadhan³

^{1,2,3}Jurusan Informatika, Universitas Jenderal Achmad Yani, Cimahi

email: sopwatunanisa20@if.unjani.ac.id^{1*}

Abstrak: *Stress* seringkali menjadi tantangan utama yang dihadapi oleh mahasiswa akibat tuntutan akademis dan sosial di lingkungan pendidikan. Faktor-faktor seperti gugup, ketidakmampuan untuk mengontrol diri, kekhawatiran, dsb. adalah beberapa pemicu *stress* yang semuanya dapat berdampak negatif terhadap kesehatan fisik dan mental mahasiswa. Penelitian ini bertujuan untuk mengidentifikasi tingkat *stress* yang dialami oleh mahasiswa dengan menggunakan metode K-Nearest Neighbors (KNN) dan mengevaluasi keakuratan hasil penelitian ini. Metode KNN digunakan untuk mengklasifikasikan tingkat *stress* mahasiswa berdasarkan kemiripan atau kedekatan dengan data lain dalam dataset. Dengan menggunakan data yang diambil dari situs data.world, hasil penelitian ini menunjukkan bahwa metode KNN mampu mencapai akurasi sebesar 91,58%. Selain itu, nilai presisi, recall, dan f1-score masing-masing adalah 76,10%, 73,11%, dan 74,17%. Penelitian ini memberikan kontribusi penting dalam memahami tingkat *stress* mahasiswa dan menunjukkan efektivitas metode KNN dalam mengklasifikasikan data *stress*. Hasil ini diharapkan dapat membantu dalam pengembangan strategi yang lebih baik untuk mengelola dan mengurangi *stress* di kalangan mahasiswa..

Kata Kunci : Klasifikasi, Metode K-Nearest Neighbors (Knn), Stress

Abstract: *Stress* is often the main challenge faced by students due to academic and social demands in the educational environment. Factors such as nervousness, inability to control oneself, worry, etc. are several stress triggers, all of which can have a negative impact on students' physical and mental health. This research aims to identify the level of stress experienced by students using the K-Nearest Neighbors (KNN) method and evaluate the accuracy of the results of this research. The KNN method is used to classify student stress levels based on similarity or closeness to other data in the dataset. By using data taken from the data.world site, the results of this research show that the KNN method is able to achieve an accuracy of 91.58%. In addition, the precision, recall, and f1-score values are 76.10%, 73.11%, and 74.17% respectively. This research makes an important contribution in understanding student stress levels and shows the effectiveness of the KNN method in classifying stress data. It is hoped that these results will help in the development of better strategies for managing and reducing stress among college students.

Keywords : Classification, K-Nearest Neighbors (Knn) Method, Stress

PENDAHULUAN

Gaya hidup modern zaman sekarang ini menyebabkan berbagai jenis gangguan kesehatan mental seperti kecemasan, *stress* dan juga depresi. *Stress* yang berkepanjangan bisa memberikan efek negatif yang mempengaruhi kesehatan fisik dan mental seseorang [1]. Gangguan kesehatan tersebut umum terjadi dan dapat memiliki dampak serius jika tidak diobati dengan cepat [2]. Tingkat *stress* yang tinggi dapat mengganggu kemampuan mahasiswa untuk belajar, berpikir secara kritis [3]. Terdapat beberapa faktor yang dapat memicu seseorang dan membuat mereka menjadi cemas dan *stress* dalam waktu lama misalnya perasaan tidak berharga, benci pada diri sendiri, putus asa [4], tekanan akademis, tekanan sosial, kurangnya dukungan dari orang-orang sekitar, dsb. [8]. Data mining dapat diartikan sebagai proses menemukan pola dalam data. Berdasarkan tugasnya, data mining dibagi menjadi deskripsi, estimasi, prediksi, klasifikasi, clustering, dan asosiasi [9]. Klasifikasi tingkat *stress* dapat membantu mengidentifikasi mahasiswa yang berisiko mengalami masalah kesehatan akibat *stress*. Klasifikasi ini merupakan salah satu teknik penting dalam pembelajaran mesin yang bertujuan untuk memprediksi kelas atau label dari data baru berdasarkan pola yang ditemukan dalam data pelatihan [10].

Berdasarkan penelitian terdahulu, mengetahui tingkat *stress* sangat penting karena hasil penelitian menunjukkan bahwa *stress* dan kecemasan dapat diidentifikasi dengan akurat menggunakan fitur fisiologis, yang dapat digunakan untuk mengembangkan platform untuk mendukung mahasiswa dalam mengelola tingkat *stress* dan kecemasan mereka [3]. Sebuah penelitian terdahulu telah mengungkap potensi luar biasa dalam mengidentifikasi tingkat *stress* menggunakan metode K-Nearest Neighbors (KNN). Dengan memanfaatkan data berharga dari kuesioner pasien yang diperoleh langsung dari Dinas Kesehatan, peneliti melakukan eksperimen dengan berbagai nilai k (5, 8, 10, dan 15). Hasilnya tingkat akurasi mencapai 82% [1]. Pada penelitian algoritma Support Vector Machine memiliki akurasi tertinggi yaitu sebesar 85,71% yang dilakukan di Jaypee Institute of Information Technology, Noida, India. Dimana penelitian ini berfokus pada membaca sinyal otak sebagai pengukuran standar stress, dan menekankan pentingnya deteksi dini *stress* untuk mencegah dampak negatif seperti depresi dan bunuh diri [5].

Penelitian mengenai klasifikasi dan deteksi tingkat *stress* telah banyak dilakukan oleh para peneliti sebelumnya salah satunya membahas penggunaan algoritma pembelajaran mesin untuk mendeteksi dan mengklasifikasi masalah kesehatan mental, khususnya dengan fokus pada tingkat kecemasan kalangan mahasiswa. Hasil dari klasifikasi *stress* dalam penelitian tersebut menunjukkan bahwa algoritma *random forest* memberikan akurasi tertinggi yaitu sebesar 78,9% dibandingkan dengan algoritma *Naïve Bayes* 71,05%, *Decision Tree* sebesar 71,05%, dan *Support Vector Machine (SVM)* yang sebesar 75,5% [4]. Pada penelitian klasifikasi diabetes hasil akhir penelitian ini memiliki akurasi tertinggi 39% pada $K=3$, presisi tertinggi 65% pada $K=3$ dan $K=5$, *recall* tertinggi 36% pada $K=3$, dan *F-Measure* tertinggi 46% pada $K=3$ [16]. Adapula penelitian dengan menggunakan kasus kinerja pemerintah di twitter dengan menggunakan rasio 80/20 dengan $k=3$ memiliki akurasi 90% [17]. Sedangkan pada klasifikasi citra makanan hanya menghasilkan akurasi sebesar 60% [10]. Hal ini tentu saja membuat semakin banyak celah untuk bisa menggunakan metode KNN.

Adapun pada penelitian lain yang membahas *Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data* [6] Penelitian ini menggunakan dataset WESAD, dengan hasil penelitian menunjukkan bahwa model yang digunakan mencapai akurasi tinggi baik dalam masalah klasifikasi tiga kelas [6]. Begitupun dengan salah satu penelitian perpaduan cerdas antara metode KNN dan teknik seleksi fitur *Symmetrical Uncertainty* telah terbukti ampuh dalam meningkatkan ketepatan klasifikasi. Inovasi ini berhasil menaikkan akurasi sebesar 3.00%, mendemonstrasikan potensi besar dalam mengoptimalkan kinerja algoritma KNN [12].

Pemanfaatan metode *K-Nearest Neighbors (KNN)* dapat menjadi sarana efektif untuk mengkategorikan dan menganalisis tingkat *stress* mahasiswa. *K-Nearest Neighbors* atau biasa disebut dengan KNN merupakan salah satu metode dalam algoritma pembelajaran mesin yang termasuk dalam kelompok *instance-based learning*. Ini juga diklasifikasikan sebagai teknik *lazy learning* [11]. Melalui data hasil dari mencari “k” tetangga terdekat dari suatu titik data yang belum diketahui kelasnya. Maka, nantinya kelas mayoritas dari k tetangga ini kemudian diatributkan kepada titik data tersebut. Misal, ketika kita memasukkan data baru, kemudian kita olah dengan KNN sehingga dapat menentukan data yang baru tersebut masuk ke dalam kelas kategori tidak *stress (no)*, *stress* ringan (*mild*), *stress* sedang (*moderate*) dan *stress* berat (*moderate*).

Dengan banyaknya penelitian untuk mengklasifikasikan tingkat *stress* menggunakan metode *machine learning*. Maka dari itu pada penelitian yang dilakukan ini, peneliti membuat klasifikasi data tingkat *stress* mahasiswa untuk menentukan tingkat *stress* mahasiswa dengan menggunakan metode KNN. Penelitian ini bertujuan untuk mengetahui sejauh mana tingkat *stress* mahasiswa di lingkungan perguruan tinggi dengan mengklasifikasikan tingkat *stress* mahasiswa dan menentukan sejauh mana metode KNN yang digunakan dapat memberikan hasil klasifikasi yang akurat berdasarkan data yang dikumpulkan.

TINJAUAN PUSTAKA

Dalam penelitian terdahulu [1] yang dijadikan sebagai acuan dalam penelitian ini terdapat kelebihan serta kekurangan. Hal ini dapat dilihat dimana dalam hasil implementasi penelitiannya menunjukkan bahwa akurasi yang didapat dengan nilai $k=4$ menghasilkan akurasi terbaik sebesar 82% dengan jumlah data latihnya sebanyak 50 data. Penelitian ini pun melakukan analisis mendalam tentang penggunaan variasi nilai k. Tetapi dengan variasi nilai k yang beragam, penelitian menyebutkan bahwa tingkat akurasi semakin menurun dengan nilai k yang lebih tinggi dan jumlah data latih yang lebih rendah. Hal ini, karena data yang digunakan relatif kecil yakni hanya 100 data. Oleh karena itu pada penelitian ini menggunakan data dari situs data.world sebesar 16.385 data dengan pemilihan nilai $k=1$. Adapun beberapa teori yang digunakan dalam penelitian ini yaitu sebagai berikut:

Stress

Stress adalah respon fisiologis dan psikologis yang terjadi ketika individu merasa tidak mampu menangani tuntutan atau tekanan tertentu. Stres terdapat beberapa macam seperti stres akademik yang terjadi saat seseorang merasa tertekan oleh tuntutan belajar, ujian, atau tugas sekolah. Stres finansial yang dialami ketika seseorang khawatir tentang keuangan mereka, seperti masalah utang atau ketidakpastian penghasilan. Dan stres pekerjaan yang muncul akibat tekanan di tempat kerja, seperti beban kerja yang berlebihan, tenggat waktu ketat, atau konflik dengan rekan kerja [3].

Data Mining

Data mining adalah proses menggunakan statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi berguna dari data. Data mining dapat diartikan sebagai proses menemukan pola dalam data. Berdasarkan tugasnya, data mining dibagi menjadi deskripsi, estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi. Proses data mining terdiri dari tiga langkah utama, yaitu persiapan data, dimana data dipilih, dibersihkan, dan diproses dari awal sesuai pedoman dan pengetahuan [9].

Klasifikasi

Klasifikasi adalah proses pengelompokkan dua atau lebih data sesuai dengan kesamaan suatu kriteria tertentu. Klasifikasi ini merupakan salah satu teknik penting dalam pembelajaran mesin yang bertujuan untuk memprediksi kelas atau label dari data baru berdasarkan pola yang ditemukan dalam data pelatihan [10].

K-Nearest Neighbors

K-Nearest Neighbor (K-NN) adalah algoritma pembelajaran mesin yang termasuk dalam kelompok *instance-based learning*. Ini juga diklasifikasikan sebagai teknik *lazy learning* [11]. Dalam metode KNN terdapat beberapa tahapan yang perlu dilakukan seperti menentukan parameter k , menghitung kedekatan berdasarkan model jarak *euclidean*, mengurutkan hasil jarak, menghitung jumlah setiap kelas berdasarkan nilai k , dan kelas mayoritas dijadikan sebagai kelas bagi data uji [12]. KNN bekerja dengan mencari kelompok k tetangga terdekat atau objek terdekat dalam data pelatihan yang paling mirip dengan objek yang diberikan dalam data baru atau data uji. Terdekat dalam konteks ini diukur berdasarkan jarak antara fitur-fitur atau atribut dari objek-objek tersebut. Jarak ini sering kali dihitung menggunakan metrik seperti jarak *Euclidean*, atau metrik lainnya. Berikut adalah rumus yang digunakan untuk menghitung jarak *euclidean* dalam penelitian ini [13]:

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

Keterangan :

d : jarak kedekatan

x : data training (data latih)

y : data testing (data uji)

n : jumlah atribut antara 1 s.d n

i : atribut individu antara 1 sampai dengan n

Confusion Matrix

Confusion matrix adalah metode untuk mengukur performa klasifikasi. Matrix ini membandingkan hasil klasifikasi sistem dengan hasil yang seharusnya. Dalam *confusion matrix*, terdapat empat istilah utama yaitu: *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* [14]. Berikut merupakan tabel 1 *Confusion Matrix* [15]:

Actual	Tabel 1. Confusion Matrix	
	Prediksi	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Keterangan:

- *True Positive (TP)*: Kondisi dimana model mengidentifikasi dengan benar data positif.
- *True Negative (TN)*: Kondisi dimana model mengidentifikasi dengan benar data negatif.
- *False Positive (FP)*: Kondisi dimana model salah mengidentifikasi data negatif sebagai positif.
- *False Negative (FN)*: Kondisi dimana model salah mengidentifikasi data positif sebagai negatif.

Dengan nilai *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* maka dapat diperoleh nilai akurasi, presisi, *recall* dan *F1 Score*.

Akurasi

Akurasi adalah rasio antara jumlah prediksi yang benar (baik *True Positive* maupun *True Negative*) dengan total jumlah data. Adapun rumus akurasi yaitu [16]:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

Presisi

Presisi adalah rasio antara jumlah prediksi positif yang benar (*True Positive*) dengan total jumlah prediksi positif (*True Positive* + *False Positive*). Dengan rumus presisi sebagai berikut [16]:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Recall

Recall adalah rasio antara jumlah prediksi positif yang benar (*True Positive*) dengan total jumlah data yang sebenarnya positif (*True Positive* + *False Negative*). Adapun rumusnya yaitu [16]:

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

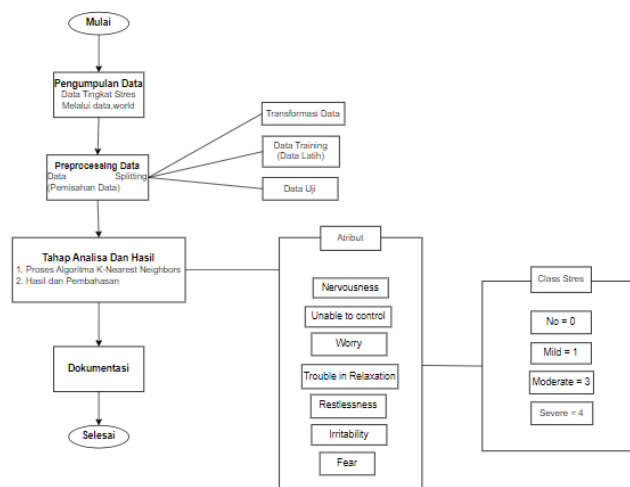
F1 Score

F1 Score adalah harmonisasi rata-rata antara presisi dan *recall*, memberikan ukuran keseimbangan antara keduanya. Dengan rumus sebagai berikut [16]:

$$F1\ Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (3)$$

METODE

Algoritma *K-Nearest Neighbors* (KNN) adalah metode yang digunakan dalam penelitian ini untuk mengklasifikasikan tingkat stress pada mahasiswa. KNN bekerja dengan cara mengklasifikasikan data baru berdasarkan identifikasi k objek terdekat (tetangga) dalam data pembelajaran. Kelas yang paling banyak dimiliki oleh k tetangga terdekat tersebut akan menjadi kelas bagi objek baru [17]. Jarak *Euclidean* biasanya digunakan untuk menghitung jarak antara objek-objek tersebut [7]. Penelitian ini dimulai dengan mengumpulkan data tentang tingkat *stress* dari sumber *data.world*. Setelah data terkumpul, dilakukan *preprocessing* yang melibatkan pemisahan data menjadi data latih dan data uji, serta transformasi data untuk mempersiapkannya untuk analisis lebih lanjut. Selanjutnya, pada tahap analisa dan hasil, data yang telah diproses digunakan dalam algoritma *K-Nearest Neighbors* (KNN) untuk mengklasifikasikan tingkat *stress* mahasiswa berdasarkan atribut seperti *nervousness*, *unable to control*, *worry*, *trouble in relaxation*, *restlessness*, *irritability*, dan *fear*. Hasil dari klasifikasi ini akan menunjukkan kelas *stress* yang terdiri dari empat kategori: tidak *stress* (0), *stress* ringan (1), *stress* sedang (2), dan *stress* berat (3). Tahap terakhir adalah dokumentasi, yang mencakup pencatatan seluruh proses dan hasil penelitian sebelum menyelesaikan penelitian. Berikut merupakan gambar alur penelitian KNN terdapat pada gambar 1.



Gambar 1. Alur metode penelitian

Pengumpulan Data

Data yang digunakan dalam penelitian klasifikasi tingkat stress mahasiswa ini adalah data yang diambil secara online dari internet melalui situs *data.world* dengan judul “*stress analysis*” dengan jumlah data sebanyak 16.385 data dan 7 atribut dengan *class* yang terdiri dari 4 kategori.

Preprocessing Data

	Nervousness	unable to control	Worry	trouble in Relaxation	Restlessness	Irritability	Fear	stress
0	Severe	Moderate	Mild	No	Mild	Mild	No	Mild
1	No	No	Mild	Mild	No	Mild	Moderate	Mild
2	Severe	Severe	Mild	Severe	Severe	Mild	Severe	Severe
3	No	Severe	Severe	Mild	Severe	No	Severe	Moderate
4	Mild	No	No	No	No	No	Mild	No
...
16379	Mild	Severe	Mild	No	Moderate	No	Mild	Mild
16380	Mild	No	No	No	Severe	No	No	No
16381	No	Moderate	Moderate	Mild	Severe	No	Moderate	Mild
16382	Severe	Mild	Severe	Mild	No	Mild	Severe	Moderate
16383	Moderate	Moderate	Moderate	Moderate	Severe	Moderate	Moderate	Moderate

16384 rows × 9 columns

Gambar 2. Tampilan Dataset

Sebelumnya dataset yang digunakan dalam penelitian ini sebanyak 16.385 data dengan atribut yang digunakan sebanyak 7 atribut sebagaimana dalam tabel 2:

Table 2. Tabel Atribut

No.	Atribut Yang Digunakan
1.	Nervousness
2.	Unable to control
3.	Worry
4.	Trouble in Relaxation
5.	Restlessness
6.	Irritability
7.	Fear

Dalam penelitian ini, berbagai perasaan dan kondisi diukur untuk menentukan tingkat *stress* mahasiswa. Pertama, kegugupan (*nervousness*) adalah perasaan cemas yang diukur dalam skala tidak ada (*no*), ringan (*mild*), sedang (*moderate*), dan berat (*severe*). Kemampuan mengontrol (*unable to control*) juga diukur dengan skala yang sama. Kekhawatiran (*worry*) adalah perasaan khawatir berlebihan yang diukur dengan skala tersebut. Kesulitan dalam relaksasi (*trouble in relaxation*) mengukur bagaimana sulitnya seseorang untuk bersantai, juga dalam skala yang sama. Kegelisahan (*restlessness*) adalah perasaan gelisah yang diukur dengan skala *no*, *mild*, *moderate*, dan *severe*. Iritabilitas (*irritability*) mengukur seberapa mudah seseorang marah atau tersinggung, menggunakan skala yang sama. Terakhir, ketakutan (*fear*) adalah perasaan takut atau cemas tanpa alasan yang jelas, juga diukur dalam skala tidak ada, ringan, sedang, dan berat. Semua variabel ini membantu dalam menentukan tingkat *stress* mahasiswa.

Transformasi data

Transformasi data (mengubah data) kategori atau label menjadi nilai numerik yang dapat digunakan oleh banyak algoritma pembelajaran mesin adalah proses yang sangat penting yang dikenal sebagai transformasi data dari bentuk kategorikal menjadi bentuk numerik. Salah satu cara untuk mengubah data kategori menjadi bentuk numerik adalah dengan mengkodekan label. Setiap kategori khusus fitur dalam *Label Encoding* diubah menjadi nilai integer yang berbeda. Dan untuk label yang digunakan dalam dataset ini terdiri dari 4 kategori kelas yaitu, *no*, *mild*, *moderate* dan *severe*.

Table 3. Tabel Tranformasi Data

Data Kategorikal	Transformasi data
No	0
Mild	1
Moderate	2
Severe	3

Pembagian data latih dan data uji

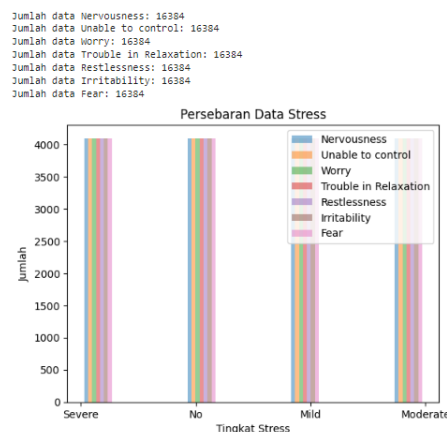
Setelah dilakukan transformasi data, kemudian dataset yang ada dibagi menjadi data latih dan data uji dengan persentase 90/10 dimana 90% untuk data latih dan 10% untuk data uji. Berikut data latih dan uji dapat dilihat dalam tabel 4 dibawah ini.

Table 4. Pembagian data latih dan data uji

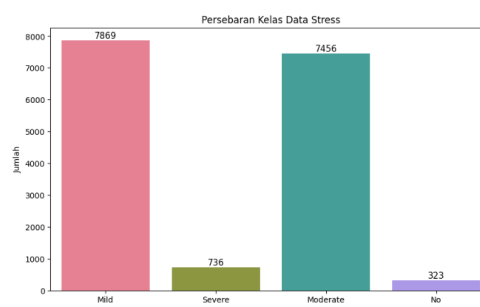
Data Latih	Data Uji
11.468	4.916

Analisis Data

Sistem klasifikasi tingkat *stress* mahasiswa menggunakan metode *K-Nearest Neighbors (K-NN)* untuk menentukan tingkat *stress* berdasarkan data yang diperoleh dari situs *data.world* dengan judul "*stress analysis*" yang terdiri dari 16.385 data. Proses dimulai dengan pengumpulan data terlebih dahulu yang kemudian dilanjutkan dengan *preprocessing* data, termasuk pembersihan, perubahan seperti dari data kategorikal menjadi data numerik, dan penyiapan data agar sesuai untuk analisis, dan pembagian data menjadi data latih dan data uji. Setelah data siap, metode K-NN diterapkan untuk mengklasifikasikan tingkat *stress* mahasiswa dengan menghitung jarak *Euclidean* antara data baru dan data yang sudah ada, dan menentukan kategori tingkat *stress* berdasarkan k tetangga terdekat. Kategori yang dihasilkan mencakup tidak *stress* (*no*), *stress* ringan (*mild*), *stress* sedang (*moderate*), dan *stress* berat (*severe*). Hasil klasifikasi dievaluasi menggunakan *confusion matrix* untuk mengukur performa sistem dengan membandingkan hasil prediksi dengan kondisi sebenarnya. Data ini kemudian dibagi menjadi data latih dan data uji untuk proses klasifikasi. Berikut persebaran data tingkat *stress* dapat pada gambar 2 dan 3.



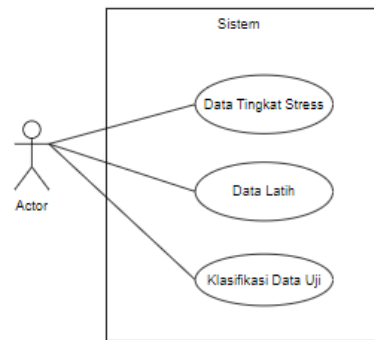
Gambar 3. Persebaran Data Stress



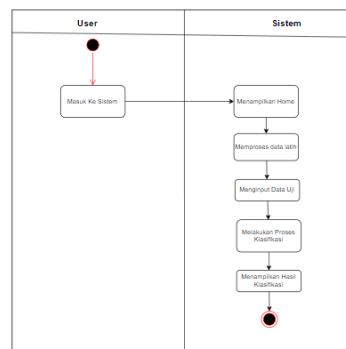
Gambar 4. Persebaran Kelas Data Stress

Perancangan Sistem

Perancangan sistem ini menggambarkan bagaimana sistem yang dibangun berfokus pada arsitektur, komponen, dan hubungan antar bagian dalam sistem perangkat lunak atau perangkat keras, dengan tujuan memastikan sistem berfungsi secara efisien. Seperti dengan membuat usecase diagram dan activity diagram. Dapat Dilihat pada gambar 5 dan gambar 6.



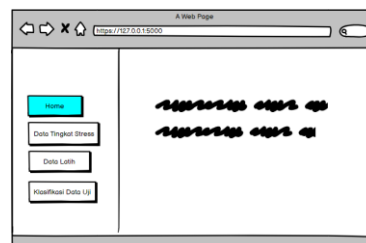
Gambar 5. Usecase Diagram



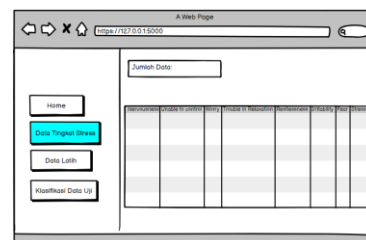
Gambar 6. Activity Diagram

Perancangan antarmuka

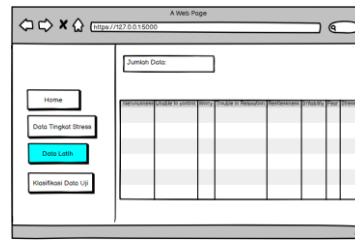
Perancangan Antarmuka Sistem adalah proses merancang tampilan dan interaksi yang digunakan oleh pengguna untuk berinteraksi dengan sistem. Tujuannya adalah untuk menciptakan antarmuka yang intuitif, menarik, dan mudah digunakan. Untuk Tampilan perancangan antarmuka dapat dilihat pada gambar 7-10.



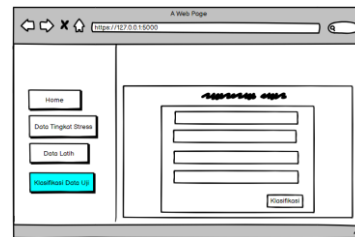
Gambar 7. Antarmuka Home



Gambar 8. Antarmuka Dataset Tingkat Stress



Gambar 9. Antarmuka Data Latih



Gambar 10. Antarmuka Klasifikasi Data Uji

HASIL DAN PEMBAHASAN

Pengujian Model Klasifikasi dengan metode KNN menggunakan Confusion Matrix

Dalam hal ini, tahapan pengujian akurasi metode *K-Nearest Neighbors* dilakukan dengan *Confusion Matrix*. Tujuan dari pengujian ini adalah untuk menentukan seberapa akurat klasifikasi data tingkat *stress* mahasiswa yang telah dibuat dengan model KNN. Pengujian ini dilakukan dengan menggunakan *confusion matrix*, dimana matrix klasifikasi dibandingkan dengan kelas data asli dari data inputan. Pengujian akurasi ini dilakukan dengan 4.916 Record dari data uji dengan nilai $k=1$.

Table 5. Tabel Hasil Confusion Matrix

True Positives (TP)	[28 729 569 38]
False Positives (FP)	[16 162 83 15]
True Negatives (TN)	[1583 681 824 1556]
False Negatives (FN)	[13 68 164 31]

True Positives (TP): [28 729 569 38] = 1.364

False Positives (FP): [16 162 83 15] = 276

True Negatives (TN): [1583 681 824 1556] = 4.644

False Negatives (FN): [13 68 164 31] = 276

Perhitungan Akurasi

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + FP + TN + FN} \quad (2) \\
 &= \frac{1.364 + 4.644}{1.364 + 276 + 4.644 + 276} \\
 &= \frac{6.008}{6.560} = 0,9158 \text{ Atau } 91,58\%
 \end{aligned}$$

Perhitungan Presisi

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Presisi 1} = \frac{28}{28+16} = \frac{28}{44} = 0,6364$$

$$\text{Presisi 2} = \frac{729}{729+162} = \frac{729}{891} = 0,8182$$

$$\text{Presisi 3} = \frac{569}{569+83} = \frac{569}{652} = 0,8726$$

$$\text{Presisi 4} = \frac{38}{38+15} = \frac{38}{53} = 0,7169$$

Rata-Rata Presisi = $\frac{0,6364+0,8182+0,8726+0,7169}{4} = 0,7610$
Sehingga rata-rata presisi nya sebesar 76,10%

Perhitungan Recall

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Recall 1 = \frac{28}{28+13} = \frac{28}{41} = 0,6829$$

$$Recall 2 = \frac{729}{729+68} = \frac{729}{797} = 0,9146$$

$$Recall 3 = \frac{569}{569+164} = \frac{569}{733} = 0,7762$$

$$Recall 4 = \frac{38}{38+31} = \frac{38}{69} = 0,5507$$

$$Rata-Rata Presisi = \frac{0,6829+0,9146+0,7762+0,5507}{4} = 0,7311$$

Sehingga rata-rata recall nya sebesar 73,11%

Perhitungan F1-Score

$$F1\ Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (5)$$

$$F1-Score 1 = 2 \times \frac{(0,6364 \times 0,6829)}{(0,6364 + 0,6829)} = 0,6588$$

$$F1-Score 2 = 2 \times \frac{(0,8182 \times 0,9146)}{(0,8182 + 0,9146)} = 0,8637$$

$$F1-Score 3 = 2 \times \frac{(0,8726 \times 0,7762)}{(0,8726 + 0,7762)} = 0,8215$$

$$F1-Score 4 = 2 \times \frac{(0,7169 \times 0,5507)}{(0,7169 + 0,5507)} = 0,6229$$

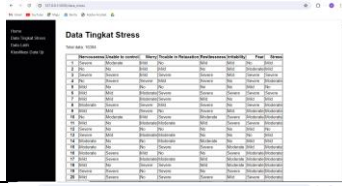
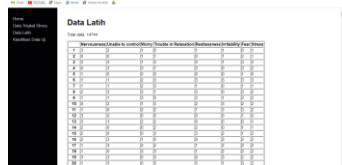
$$Rata-Rata F1-Score = \frac{0,6588+0,8637+0,8215+0,6229}{4} = 0,7417$$




Sehingga rata-rata recall nya sebesar 74,17%

Pengujian Sistem dengan Black Box Testing beserta Tampilan Implementasi Antarmuka

Pengujian adalah proses yang digunakan untuk menguji sistem untuk menemukan kesalahan. Ini juga bertujuan untuk menunjukkan bahwa fungsi-fungsi perangkat lunak sesuai dengan spesifikasinya. Pengujian *black box testing* adalah metode yang digunakan untuk pengujian ini. Pengujian *black box testing* hanya melihat hasil eksekusi melalui data uji dan memeriksa fungsionalitas sistem. Pengujian *black box testing* dilakukan dengan menjalankan atau mengeksekusi unit atau modul, dan kemudian melihat apakah hasil unit memenuhi proses yang diinginkan.

Table 6. Pengujian Sistem dengan BlackBox Testing

Nama Fungsi	Kode Uji	Kasus Uji	Hasil Yang Diharapkan	Hasil
Data Tingkat Stress	UC01	Sistem menampilkan dataset tingkat stress secara keseluruhan		Sesuai
Data Latih	UC02	Sistem menampilkan data latih yang digunakan		Sesuai

Nama Fungsi	Kode Uji	Kasus Uji	Hasil Yang Diharapkan	Hasil
Klasifikasi Data Uji	UC03	Sistem dapat menerima input dan menghitung klasifikasi dari data yang diinputkan		Sesuai
				
				

Hasil Pengujian Confusion Matrix dan Black Box Testing

Setelah melihat hasil pengujian akurasi metode *Confusion Matrix*, maka didapatkan hasil untuk mengukur kinerja atau performa algoritma *K-Nearest Neighbors*. Ini diukur dengan empat metode evaluasi *Confusion Matrix* yaitu akurasi, presisi, *recall*, dan skor *f1*. Hasil pengujian metode *Confusion Matrix* didapatkan:

Nilai Akurasi sebesar 91,58%

Nilai Presisi sebesar 76,10%

Nilai Recall sebesar 73,11%

Nilai F1-Score sebesar 74,17%

Dan untuk hasil pengujian menggunakan *blackbox testing* yaitu:

Jumlah Kode Uji = 3 Kode Uji

Kode Uji Hasil Sesuai = 3 Kode Uji

Kode Uji Hasil Tidak Sesuai = 3 Kode Uji

Persentase:

$$\begin{aligned}
 \text{Persentase} &= \frac{(\text{jumlah kode uji} - \text{kode uji tidak sesuai})}{(\text{jumlah kode uji})} \times 100\% \\
 &= \frac{(3 - 0)}{(3)} \times 100\% \\
 &= 100\%
 \end{aligned}$$

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa sistem untuk mengklasifikasikan tingkat *stress* mahasiswa menggunakan metode *K-Nearest Neighbors* berhasil diimplementasikan dengan baik. Hal ini dapat dilihat dari hasil pengujian menggunakan *confusion matrix* yang menunjukkan tingkat akurasi yang lebih besar dibandingkan dengan penelitian terdahulu sebesar 82%, sedangkan penelitian ini menghasilkan akurasi sebesar 91,58%, dengan nilai presisi, *recall*, dan *F1-score* masing-masing adalah 76,10%, 73,11%, dan 74,17%. Begitupun dengan hasil pengujian *black box testing* yang mencatat 100% kesesuaian terhadap skenario yang dirancang sebelumnya. Dengan demikian, sistem ini berhasil dalam memenuhi tujuan untuk membantu mengelola dan mengidentifikasi tingkat *stress* mahasiswa secara efektif.

Saran

Penelitian selanjutnya bisa mempertimbangkan penggunaan data tambahan yang lebih variatif dan lebih besar untuk meningkatkan generalisasi model dengan menggunakan metode *machine learning* lain, selain dari

metode KNN. Serta mengintegrasikan fitur-fitur pengujian tambahan seperti validasi silang (*cross-validation*) untuk mengevaluasi performa model secara lebih mendalam.

DAFTAR PUSTAKA

- [1] V. R. Wirayudha, N. Hidayat, and R. K. Dewi, "Identifikasi Tingkat Stress Pada Manusia Menggunakan Metode K-NN (K-Nearest Neighbour)," ... *Tekno. Inf. dan Ilmu* ..., vol. 4, no. 9, pp. 3129–3134, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/7886%0Ahttps://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/7886/3708>.
- [2] A. Daza Vergaray, J. C. H. Miranda, J. B. Cornelio, A. R. López Carranza, and C. F. Ponce Sánchez, "Predicting the depression in university students using stacking ensemble techniques over oversampling method," *Informatics Med. Unlocked*, vol. 41, no. June, 2023, doi: 10.1016/j.imu.2023.101295.
- [3] J. Rodríguez-Arce, L. Lara-Flores, O. Portillo-Rodríguez, and R. Martínez-Méndez, "Towards an anxiety and stress recognition system for academic environments based on physiological features," *Comput. Methods Programs Biomed.*, vol. 190, 2020, doi: 10.1016/j.cmpb.2020.105408.
- [4] S. Bhatnagar, J. Agarwal, and O. R. Sharma, "Detection and classification of anxiety in university students through the application of machine learning," *Procedia Comput. Sci.*, vol. 218, no. 2022, pp. 1542–1550, 2022, doi: 10.1016/j.procs.2023.01.132.
- [5] R. Ahuja and A. Banga, "Mental stress detection in university students using machine learning algorithms," *Procedia Comput. Sci.*, vol. 152, pp. 349–353, 2019, doi: 10.1016/j.procs.2019.05.007.
- [6] P. Bobade and M. Vani, "Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data," *Proc. 2nd Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2020*, no. December, pp. 51–57, 2020, doi: 10.1109/ICIRCA48905.2020.9183244.
- [7] M. Stadnyk *et al.*, "Steady state visual evoked potential classification by modified KNN method," *Procedia Comput. Sci.*, vol. 207, no. Kes, pp. 71–79, 2022, doi: 10.1016/j.procs.2022.09.039.
- [8] R. E. Demirbatir, "Undergraduate Music Student's Depression, Anxiety and Stress Levels: A Study from Turkey," *Procedia - Soc. Behav. Sci.*, vol. 46, pp. 2995–2999, 2012, doi: 10.1016/j.sbspro.2012.05.603.
- [9] F. T. Admojo and Ahsanawati, "Klasifikasi Aroma Alkohol Menggunakan Metode KNN," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 34–38, 2020, doi: 10.33096/ijodas.v1i2.12.
- [10] M. R. Setiawan, Y. A. Sari, and P. P. Adikara, "Klasifikasi Citra Makanan Menggunakan K-Nearest Neighbor dengan Fitur Bentuk Simple Morphological Shape Descriptors dan Fitur Warna Grayscale Histogram," vol. 3, no. 3, pp. 2726–2731, 2019.
- [11] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [12] A. K. B. Ginting, M. S. Lydia, and E. M. Zamzami, "Peningkatan Akurasi Metode K-Nearest Neighbor dengan Seleksi Fitur Symmetrical Uncertainty," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1714, 2021, doi: 10.30865/mib.v5i4.3254.
- [13] M. Çakir, M. Yilmaz, M. A. Oral, H. Ö. Kazanci, and O. Oral, "Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture," *J. King Saud Univ. - Sci.*, vol. 35, no. 6, 2023, doi: 10.1016/j.jksus.2023.102754.
- [14] F. Istighfarizky, N. A. Sanjaya ER, I. M. Widiartha, L. G. Astuti, I. G. N. A. C. Putra, and I. K. G. Suhartana, "Klasifikasi Jurnal menggunakan Metode KNN dengan Mengimplementasikan Perbandingan Seleksi Fitur," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 11, no. 1, p. 167, 2022, doi: 10.24843/jlk.2022.v11.i01.p18.
- [15] M. A. Imron and B. Prasetyo, "Improving Algorithm Accuracy K-Nearest Neighbor Using Z-Score Normalization and Particle Swarm Optimization to Predict Customer Churn," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 56–62, 2020, doi: 10.52465/josce.v1i1.7.
- [16] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [17] O. S. Y. Prakasa and K. M. Lhaksamana, "Klasifikasi dengan Menggunakan Algoritma K-Nearest Neighbor pada Kasus Kinerja Pemerintah di Twitter," *eProceedings Eng.*, vol. 5, no. 3, pp. 8237–8248, 2018.