



Analisis Perbandingan Algoritma *K-Nearest Neighbor* dan *Decision Tree* Pada Klasifikasi Tingkat *Stress* Individu

Vanni Manurung*, Anief Fauzan Rozi

Prodi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia

Email: ^{1,*}vannimanurung02@gmail.com, ²anief@mercubuana-yogya.ac.id

Email Penulis Korespondensi: vannimanurung02@gmail.com

Abstrak—*Stress* merupakan respons yang dialami seseorang terhadap perubahan keadaan, dan tingkat *stress* yang tinggi dapat berdampak negatif pada kesehatan fisik dan mental. Mengidentifikasi tingkat *stress* dengan tepat sangat penting untuk intervensi medis dan psikologis yang sesuai. Penelitian ini bertujuan untuk menentukan algoritma yang lebih efektif dan akurat antara *K-Nearest Neighbor* (KNN) dan *Decision Tree* dalam mengklasifikasikan tingkat *stress* individu. Data yang digunakan dalam penelitian ini diperoleh dari dataset “*Human Stress Detection*” di situs web Kaggle, yang mencakup variabel *humidity*, *temperature*, *step count*, dan *stress level*. Hasil penelitian menunjukkan bahwa algoritma *K-Nearest Neighbor* berhasil mencapai tingkat akurasi sempurna sebesar 100%, sementara *Decision Tree* mencapai akurasi 99,50%. Selain itu, dalam hal *precision*, *recall*, dan *F1-score*, *K-Nearest Neighbor* juga unggul dengan nilai masing-masing 100%, sedangkan *Decision Tree* memiliki *precision* 99,45%, *recall* 99,54%, dan *F1-score* 99,50%. Analisis juga menemukan bahwa suhu tubuh tinggi (>33°C) dan kelembapan tinggi (>22,5) berkaitan dengan tingkat *stress* lebih tinggi. Jumlah langkah di bawah 90 juga dapat mengindikasikan *stress* normal atau tinggi. Dengan demikian, penelitian ini menyimpulkan bahwa algoritma *K-Nearest Neighbor* lebih efektif dalam mengklasifikasikan tingkat *stress* individu dan faktor-faktor seperti suhu tubuh, kelembapan, dan langkah kaki memiliki peran penting dalam menentukan tingkat *stress* seseorang.

Kata Kunci: Perbandingan; KNN; Decision Tree; Klasifikasi; Tingkat Stress

Abstract—*Stress* is an individual’s response to changing circumstances, and high levels of stress can have a negative impact on physical and mental health. Correctly identifying stress levels is essential for appropriate medical and psychological interventions. This research aims to determine the more effective and accurate algorithm between *K-Nearest Neighbor* (KNN) and *Decision Tree* in classifying individual stress levels. The data used in this study was obtained from the “*Human Stress Detection*” dataset on the Kaggle website, which includes variables of humidity, temperature, step count, and stress level. The results showed that the *K-Nearest Neighbor* algorithm managed to achieve a perfect accuracy rate of 100%, while the *Decision Tree* achieved an accuracy of 99,50%. In addition, in terms of precision, recall, and F1-score, *K-Nearest Neighbor* also excelled with 100% each, while *Decision Tree* had 99,45% precision, 99,54% recall, and 99,50% F1-score. The analysis also found that high body temperature (>30°C) and high humidity (>22,5) were associated with higher stress levels. A step count below 90 can also indicate normal or high stress. Thus, this study concludes that the *K-Nearest Neighbor* algorithm is more effective in classifying an individual’s stress level and factors such as body temperature, humidity, and footsteps play an important role in determining an individual’s stress level.

Keywords: Comparison; KNN; Decision Tree; Classification; Stress Levels

1. PENDAHULUAN

Stress merupakan gejala yang dialami seseorang, baik secara fisik maupun non-fisik yang terjadi akibat dari proses penyesuaian diri terhadap perubahan keadaan yang sedang terjadi. Kondisi *stress* bisa dipicu oleh berbagai faktor, seperti aspek fisik dan lingkungan. *Stress* menjadi tantangan yang signifikan dalam kesehatan seseorang karena dapat berdampak negatif pada kesehatan fisik dan mental jika tidak ditangani dengan baik. *Stress* diklasifikasikan menjadi tiga tingkatan yaitu *stress* ringan, *stress* sedang, dan *stress* berat (Fardiana Risa et al., 2021). Menurut WHO (*World Health Organization*), lebih dari 350 juta penduduk di seluruh dunia mengalami kejadian *stress*, menempatkannya di peringkat ke-4 penyakit di dunia (Prabamurti, 2019).

Tingkat *stress* merupakan aspek krusial dalam kehidupan makhluk hidup, termasuk manusia, dan jika tingkat *stress* manusia tinggi dapat menyebabkan timbulnya penyakit (Fathirachman Mahing et al., n.d.). Tingkat *Stress* yang dialami oleh individu memerlukan intervensi medis atau psikologi. Oleh karena itu, penting untuk mengidentifikasi dan memahami tingkat *stress* individu untuk memberikan intervensi yang sesuai. Teknik-teknik klasifikasi data menjadi kunci untuk mengkategorikan *stress* berdasarkan faktor-faktor yang mempengaruhinya. Pengukuran tingkat akurasi pada dataset dapat dilakukan melalui teknik klasifikasi *data mining*, yang merupakan suatu proses penilaian data untuk menempatkannya dalam kategori tertentu dari berbagai kategori yang ada, dengan membuat pola dari data yang telah diberikan dan menggunakannya untuk mengklasifikasikan data baru. (Utomo & Mesran, 2020).

(Rachakonda et al., 2019) melakukan penelitian tentang pengukuran dan analisis tingkat *stress* dengan memanfaatkan perangkat IoMT (*Internet of Medical Things*). IoMT merupakan jaringan perangkat medis yang terhubung ke internet dan memungkinkan pengumpulan, transmisi, dan analisis data kesehatan secara *real-time*. Keunggulan utama dari IoMT adalah kemampuannya untuk memantau berbagai aspek kesehatan individu secara terus menerus dan terintegrasi, memberikan data yang kaya dan mendalam yang dapat digunakan untuk mengidentifikasi kondisi kesehatan seperti *stress*. Dalam penelitian ini, berbagai parameter fisiologi dan perilaku diukur untuk memberikan gambaran komprehensif tentang kondisi *stress* seseorang. Penelitian ini bertujuan untuk mengembangkan sensor cerdas untuk mendeteksi *stress* pada individu menggunakan IoMT. Hasil penelitian menunjukkan bahwa sensor cerdas ini memiliki kemampuan dengan hasil akurasi 97%. Hasil penelitian juga menunjukkan bahwa kelembapan

udara, suhu tubuh, dan jumlah langkah harian dapat mempengaruhi tingkat *stress* seseorang. Hal ini menunjukkan bahwa IoMT dapat digunakan untuk mendeteksi *stress* dan membantu dalam pengawasan kesehatan mental dan fisik individu.

Penelitian yang dilakukan oleh (Van Fadhila et al., 2023) bertujuan untuk membantu dalam mendeteksi tingkat *stress* berdasarkan kualitas tidur manusia menggunakan *K-Nearest Neighbor*, *Decision Tree*, dan SVM. Hasil yang diperoleh dari penelitian ini yaitu nilai *Accuracy*, *Precision*, *Recall*, *F1-score* pada metode KNN dan SVM sebesar 100% dan pada metode *Decision Tree* menghasilkan *Accuracy* sebesar 99% dan *Precision*, *Recall*, *F1-score* sebesar 100%. Penelitian ini dapat mengevaluasi keakuratan model klasifikasi dalam mengidentifikasi individu yang diprediksi masuk dalam kategori tingkat *stress* yang telah ditetapkan.

Penelitian oleh (Wibowo et al., 2023), melakukan penelitian perbandingan teknik klasifikasi untuk mengidentifikasi *stress* pada mahasiswa di lembaga pendidikan tinggi. Penelitian ini menggunakan metode klasifikasi *Naive Bayes*, *Random Tree*, *Support Vector Machine* (SVM), *Neural Network*, *Decision Tree*, *Random Forest*, dan *K-Nearest Neighbor* (KNN) dengan memanfaatkan data mahasiswa. Metode *Random Tree* dan *Decision Tree* dipilih sebagai metode yang paling efektif untuk hasil perbandingan performa ini karena mencapai 100% dengan perbandingan 80:20.

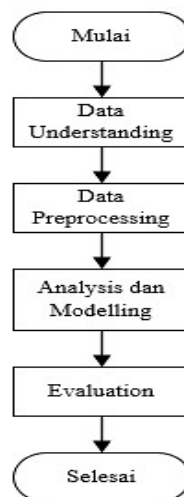
Penelitian yang dilakukan oleh (Wirayudha et al., 2020) dengan meneliti tingkat *stress* pada manusia menggunakan metode KNN menghasilkan tingkat akurasi KNN sebesar 82%. Pengujian akurasi ini dilakukan dengan menerapkan algoritma KNN menggunakan variasi nilai K (5, 8, 10, dan 15) serta beragam ukuran data latihan (8, 18, 38, dan 50) pada sebuah *dataset* yang mencatat gejala dan bobotnya pada pasien.

Penelitian lain yang dilakukan oleh (Ahmad et al., 2022) dengan meneliti hubungan tingkat *stress* pada mahasiswa angkatan 2017 Fakultas Kedokteran Universitas Baiturrahman dengan insiden insomnia yang mereka alami. Temuan penelitian menunjukkan adanya hubungan antara tingkat *stress* dengan kejadian insomnia yang dialami oleh mahasiswa.

Penelitian ini penting untuk dilakukan karena *stress* menjadi masalah kesehatan yang semakin meningkat di masyarakat modern. Dengan memahami berbagai faktor yang mempengaruhi tingkat *stress* dan menggunakan metode *data mining*, kita dapat mengembangkan metode yang lebih efektif dalam mengelola *stress*. Penelitian ini diharapkan tidak hanya memberikan wawasan baru, tetapi juga menjadi landasan untuk studi lanjutan yang dapat lebih mendalam tentang penanganan *stress* dalam konteks yang lebih luas dan berkelanjutan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Pada penelitian ini, ada beberapa langkah yang harus dilakukan untuk mencapai hasil yang optimal. Tahapan penelitian ini meliputi *data understanding*, *data preprocessing*, *analysis dan modelling*, dan *evaluation*.

2.1.1 Data Understanding

Data understanding merupakan tahap awal pengumpulan data, pengecekan data, dan identifikasi kualitas data (Dhewayani et al., 2022). Data yang digunakan dalam penelitian ini diperoleh dari situs web Kaggle dengan judul “*Human Stress Detection*”. *Dataset* ini terdiri dari 2001 entri data yang mencakup tiga variabel yang dapat mempengaruhi tingkat *stress* individu, yaitu *humidity*, *temperature*, dan *step count*. Selain itu, *dataset* ini juga mencakup klasifikasi tingkat *stress* yang sudah dikalibrasi, yang terdiri dari kategori *low*, *normal*, dan *high*, yang kemudian ditetapkan ke dalam bentuk biner 0, 1, 2.



Variabel-variabel ini dipilih berdasarkan penelitian sebelumnya yang menunjukkan bahwa faktor-faktor fisiologis dan aktivitas fisik sehari-hari memiliki korelasi yang signifikan dengan tingkat *stress* seseorang. *Humidity* mengukur kadar kelembapan udara yang dapat mempengaruhi kenyamanan dan tingkat dehidrasi tubuh. *Temperature* mengukur suhu tubuh yang dapat meningkat seiring dengan tingginya tingkat *stress*. *Step count* mengukur aktivitas fisik yang berperan dalam manajemen *stress* dan kesejahteraan umum. Untuk memberikan gambaran lebih jelas mengenai dataset yang digunakan, berikut adalah lima entri data teratas dari dataset tersebut. Tabel 1 dibawah ini menampilkan lima entri data teratas dari *dataset* tersebut.

Tabel 1. Lima entri data teratas

<i>Humidity</i>	<i>Temperature</i>	<i>Step count</i>	<i>Stress level</i>
21.33	90.33	123	1
21.41	90.41	93	1
27.12	96.12	196	2
27.64	96.64	177	2
10.87	79.87	87	0

2.1.2 Data Preprocessing

Preprocessing data menjadi tahapan yang penting sebelum melakukan analisis perbandingan antara algoritma KNN dan *Decision Tree* pada klasifikasi tingkat *stress*. Menurut (Hambardzumyan, 2021), *Preprocessing data* merupakan satu langkah penting dalam *data mining* yang berurusan dengan persiapan data dan transformasi *dataset*, serta bertujuan untuk membuat proses penemuan pengetahuan lebih efisien. Tahap *preprocessing data* dilakukan untuk membersihkan data, memastikan tidak ada nilai yang hilang yang dapat menghambat pemrosesan data (Tyas & Purnamasari, 2023). Proses ini melibatkan beberapa langkah penting untuk memastikan kebersihan dan kualitas *dataset*. Tahapan pertama dalam *preprocessing data* adalah identifikasi dan penanganan nilai yang hilang (*missing values*) dalam *dataset*. Setelah memastikan tidak ada nilai yang hilang, selanjutnya dilakukan pengecekan data yang duplikat untuk menghindari bias dalam hasil analisis. Langkah selanjutnya adalah deteksi *outlier* menggunakan metode *Interquartile Range (IQR)*. Dengan menggunakan *IQR*, setiap variabel diperiksa guna memastikan tidak ada *outlier* yang signifikan. Selanjutnya dilakukan konversi satuan suhu yang awalnya dalam satuan *Fahrenheit* diubah menjadi satuan *Celsius*. Setelah itu, dilakukan analisis korelasi untuk mengidentifikasi sejauh mana variabel-variabel seperti kelembapan, suhu tubuh, dan jumlah langkah berkorelasi dengan tingkat *stress*. Terakhir, dilakukan pemeriksaan persebaran kelas target untuk memastikan setiap kelas tingkat *stress* memiliki jumlah data yang seimbang. Langkah-langkah tersebut penting untuk memastikan integritas dan konsistensi data sebelum dilakukan proses klasifikasi. Dengan demikian, untuk mengatasi masalah tersebut, perlu dilakukan *preprocessing* sebagai langkah penting untuk menghilangkan gangguan yang dapat mempengaruhi hasil klasifikasi (Purbolaksono et al., 2021).

2.1.3 Analisis dan Modelling

Langkah awal dalam *analysis* dan *modelling* adalah transformasi data. Kolom '*Stress Level*' yang awalnya numerik diubah menjadi kategori dengan nama '*Stress Level Categorical*', dimana nilai-nilai numerik 0,1, dan 2 masing-masing diubah menjadi '*low*', '*normal*', dan '*high*' untuk mempermudah interpretasi. Setelah itu, dilakukan visualisasi data menggunakan *pairplot* dari *Seaborn* untuk mengeksplorasi hubungan antar variabel terkait tingkat *stress*. Variabel-variabel dalam *dataset* juga dinormalisasikan menggunakan *MinMaxScaler* agar semua variabel memiliki rentang yang sama.

Pada tahap ini, dilakukan perbandingan kinerja antara algoritma KNN dan *Decision Tree* dalam klasifikasi tingkat *stress*. Perbandingan dilakukan untuk menentukan algoritma mana yang lebih efisien dan akurat dalam mengklasifikasikan tingkat *stress* berdasarkan *dataset* yang digunakan.

2.1.3.1 K-Nearest Neighbor

K-Nearest Neighbor (KNN) merupakan algoritma *supervised learning* dimana klasifikasi *instance* baru bergantung pada mayoritas kategori dalam KNN (Cholil et al., 2021). Algoritma ini bertujuan untuk mengklasifikasikan objek baru dengan mempertimbangkan atribut dan kedekatan dari data latih. Menurut (Suntoro, 2018), tahapan langkah dalam algoritma KNN adalah sebagai berikut:

1. Memilih nilai parameter k: Parameter k adalah jumlah tetangga terdekat yang akan dipertimbangkan dalam proses klasifikasi.
2. Melakukan perhitungan jarak *euclidean* kuadrat antara objek dan *dataset* latih yang tersedia: Jarak *euclidean* digunakan untuk mengukur kedekatan antara data baru dan data yang sudah ada dalam *dataset*.
3. Mengurutkan hasil poin secara *ascending* (dari nilai tinggi ke rendah): Setelah menghitung jarak, semua poin diurutkan berdasarkan jarak terdekat.
4. Mengidentifikasi kategori: Kategori yang dipilih adalah yang memiliki jumlah nilai k tertinggi dalam data uji. Dengan kata lain, *instance* baru diklasifikasikan ke dalam kelas yang paling umum di antara k tetangga terdekat.



Algoritma KNN memanfaatkan *neighborhood classification* sebagai nilai prediksi untuk *instance* baru. Hal ini berarti bahwa prediksi dibuat berdasarkan kedekatan dengan data yang ada, sehingga semakin banyak data yang mirip di sekitar *instance* baru, semakin akurat prediksinya.

2.1.3.2 Decision Tree

Decision Tree merupakan struktur sekuensial yang dimulai dari akar, kemudian mengevaluasi fitur dan memilih salah satu dari dua cabang, terus berlanjut hingga mencapai cabang terakhir yang disebut daun, yang umumnya merupakan representasi dari target yang dicari (Haganta Depari et al., n.d.). Umumnya, simpul teratas dari pohon *Decision Tree* disebut simpul akar, yang merupakan atribut paling berpengaruh terhadap suatu kelas tertentu (Rahayu & dkk, 2024). Menurut (Larose, 2005), *Decision Tree* dibangun menggunakan langkah-langkah sebagai berikut:

1. Menetapkan atribut sebagai simpul awal: Proses dimulai dengan memilih atribut yang paling informatif untuk dijadikan simpul akar. Informasi ini biasanya diukur dengan *gain* informasi atau *gain ratio*.
2. Membuat cabang untuk setiap nilai atribut: Setiap nilai dari atribut yang dipilih akan menghasilkan cabang yang berbeda dalam pohon keputusan.
3. Memisahkan kasus ke dalam masing-masing cabang: Data dipecah ke dalam cabang-cabang berdasarkan nilai atribut.
4. Mengulangi proses untuk setiap cabang hingga semua kasus dalam cabang memilih kelas yang serupa: Proses ini berlanjut secara rekursif untuk setiap cabang menggunakan atribut yang tersisa hingga semua data dalam satu cabang memiliki kelas yang sama atau tidak ada atribut yang tersisa.

Decision Tree memanfaatkan konsep *entropy* dan *gain* informasi untuk memilih atribut terbaik yang memisahkan data secara efektif. *Entropy* mengukur ketidakpastian atau impuritas dalam data, sedangkan *gain* informasi menghitung penurunan impuritas setelah memisahkan data berdasarkan atribut tertentu.

2.1.4 Evaluation

Tahap evaluasi dilakukan untuk mengukur sejauh mana data yang dihasilkan dapat memenuhi tujuan klasifikasi (Alghifari & Juardi, 2021). Pertama-tama, dilakukan perhitungan akurasi untuk masing-masing model menggunakan data uji. Langkah ini dilakukan untuk membandingkan hasil prediksi model dengan label sebenarnya dari data uji. Hasil akurasi dari kedua model dibandingkan untuk mengetahui apakah ada perbedaan signifikan dalam kinerja keduanya. Perbandingan model ini ditampilkan dalam diagram batang yang memperlihatkan akurasi kedua model. Selanjutnya, dilakukan evaluasi pengukuran *precision*, *recall*, dan *f1-score* untuk mengetahui perbedaan kinerja secara lebih terperinci dari kedua model. Perbandingan nilai *precision*, *recall*, dan *f1-score* dari kedua model ditampilkan dalam diagram untuk memudahkan pemahaman.

3. HASIL DAN PEMBAHASAN

3.1 Data Understanding

Pada tahap ini, pengumpulan dan pemahaman data dilakukan guna mendapatkan pemahaman yang mendalam dan kualitas dari data yang tersedia bagi peneliti. *Dataset* yang digunakan dalam penelitian ini diperoleh dari situs web Kaggle yang terdiri dari 2001 entri data. *Dataset* ini mencakup berbagai fitur yang mencakup *humidity*, *temperature*, *step count*, *stress level*. Tampilan dari *dataset* yang dimanfaatkan dalam penelitian ini diperlihatkan pada Tabel 2 berikut.

Tabel 2. *Dataset Human Stress Detection*

	<i>Humidity</i>	<i>Temperature</i>	<i>Step Count</i>	<i>Stress Level</i>
1	21.33	90.33	123	1
2	21.41	90.41	93	1
3	27.12	96.12	196	2
4	27.64	96.64	177	2
5	10.87	79.87	87	0
...
1997	21.82	90.82	96	1
1998	10.45	79.45	45	0
1999	27.22	96.22	135	2
2000	12.46	81.46	64	0
2001	16.87	85.87	50	1

3.2 Data Preprocessing

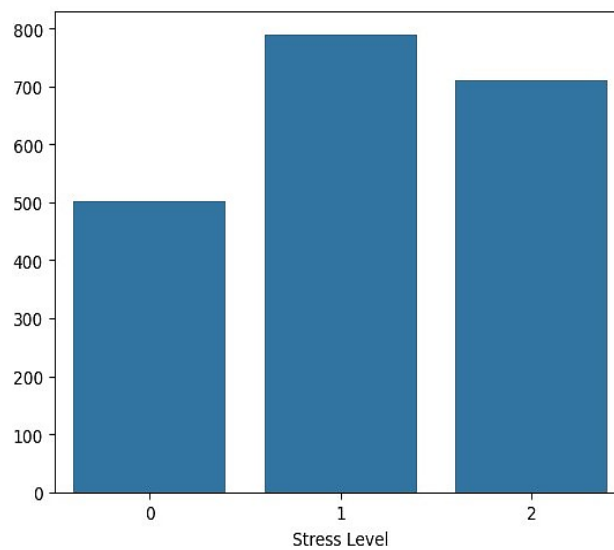
Data preprocessing bertujuan untuk memastikan bahwa data yang akan dianalisis memiliki kualitas yang baik dan siap digunakan dalam model klasifikasi. Proses ini dimulai dari pengecekan nilai yang hilang (*missing values*) dan hasilnya menunjukkan bahwa tidak ada nilai yang hilang dalam *dataset*.

Tabel 3. Jumlah data yang kosong pada *dataset*

Jumlah Data yang kosong	Jumlah
Humidity	0
Temperature	0
Step Count	0
Stress Level	0

Selanjutnya dilakukan pengecekan data duplikat yang menunjukkan bahwa tidak ada entri duplikat, memastikan bahwa setiap data point adalah unik. Deteksi *outlier* dilakukan menggunakan metode *Interquartile Range* (IQR), namun tidak ditemukan adanya *outlier* pada semua variabel, yang menunjukkan bahwa data bersih dari nilai ekstrim. Suhu yang awalnya dalam satuan *Fahrenheit* kemudian dikonversi menjadi *Celcius* untuk kemudahan interpretasi. Analisis korelasi menggunakan *heatmap* menunjukkan bahwa semua fitur memiliki korelasi yang signifikan dengan tingkat *stress*, sehingga semua fitur dalam *dataset* digunakan untuk analisis lebih lanjut.

Pada Gambar 2 menunjukkan persebaran kelas target yang ditampilkan dengan plot jumlah data per kategori tingkat *stress*, yang menunjukkan bahwa setiap kelas memiliki jumlah data yang hampir sama, sehingga tidak perlu resampling.

**Gambar 2.** Persebaran kelas target

Tahap preparation ini memastikan bahwa *dataset* yang digunakan siap untuk proses klasifikasi menggunakan algoritma KNN dan *Decision Tree*, dengan hasil analisis yang akurat dan handal.

3.3 Analysis dan Modelling

Tahap *Analysis* dan *Modelling* dimulai dengan mempersiapkan data untuk analisis lebih lanjut. Langkah pertama adalah mengubah kolom '*Stress Level*' yang berbentuk numerik ke bentuk kategori dengan menambahkan kolom baru '*Stress level categorical*'. Pada kolom ini, nilai 0 diubah menjadi '*low*', nilai 1 diubah menjadi '*normal*', dan nilai 2 menjadi '*high*'. Setelah itu, dilakukan visualisasi data menggunakan *pairplot* dari *Seaborn* untuk menganalisis hubungan antara variabel-variabel dalam *dataset* berdasarkan kelas *stress level* yang berbeda. *Pairplot* ini memperlihatkan hubungan antara *temperature* dan *humidity* dengan tingkat *stress*, menunjukkan bahwa semakin tinggi nilai *temperature* dan *humidity*, semakin tinggi pula tingkat *stress*.

Selain itu, ditemukan bahwa tingkat *stress* lebih dipengaruhi oleh *humidity* daripada *step count* atau *temperature*. Saat *step count* rendah, tingkat *stress* dipengaruhi oleh *humidity*, dimana tingkat *stress* rendah saat *humidity* rendah, dan *normal* saat *humidity* sedang. Sehingga, visualisasi *pairplot* membantu dalam memahami hubungan variabel dan tingkat *stress*, penting dalam pengembangan model klasifikasi tingkat *stress* yang akurat.

Plot bar dari rata-rata *humidity* dan *temperature* pada setiap tingkat *stress* memperkuat penelitian ini. Sebaliknya, hubungan antara *step count* dan *stress level* menunjukkan bahwa pada *step count* rendah (kurang dari 90), terdapat dua kelas *stress level* (*low* dan *normal*), mengindikasikan bahwa ada faktor lain yang mempengaruhi *stress* pada *step count* rendah. Namun, pada *step count* lebih dari 90, *stress level* cenderung berada pada *level high*. Setelah analisis awal ini, fitur-fitur *dataset* kemudian dinormalisasi menggunakan *MinMaxScaler* untuk memastikan bahwa semua fitur berada dalam rentang yang sama, sehingga model yang digunakan nantinya dapat bekerja lebih efisien.

Selanjutnya dilakukan pelatihan model KNN menggunakan parameter terbaik yang diperoleh melalui *GridSearchCV*, dan hasil prediksi diuji menggunakan data uji. *Classification report* menunjukkan bahwa model KNN memiliki akurasi yang sangat tinggi dengan *precision*, *recall*, dan *F1-score* 1.0 untuk semua kelas. Hasil yang diperoleh dapat dilihat pada Tabel 4 dibawah ini.

Tabel 4. Laporan klasifikasi model KNN

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Low</i>	1.00	1.00	1.00	107
<i>Normal</i>	1.00	1.00	1.00	142
<i>High</i>	1.00	1.00	1.00	152
<i>Accuracy</i>			1.00	401
<i>Macro avg</i>	1.00	1.00	1.00	401
<i>Weighted avg</i>	1.00	1.00	1.00	401

Setelah itu, dilakukan juga pelatihan model *Decision Tree* menggunakan parameter yang telah diperoleh. Hasil pada Tabel 5 menunjukkan kinerja model *Decision Tree* juga sangat baik. Setiap kelas dengan tingkat *stress* tinggi, rendah, dan normal, memiliki nilai *precision* sebesar 1.00, 0.99, dan 0.99, dengan *f1-core* sebesar 1.00 untuk kelas tingkat *stress* tinggi dan rendah, serta 0.99 untuk kelas tingkat *stress* normal. Nilai *recall* yang sedikit kurang dari 1.00 untuk kelas normal.

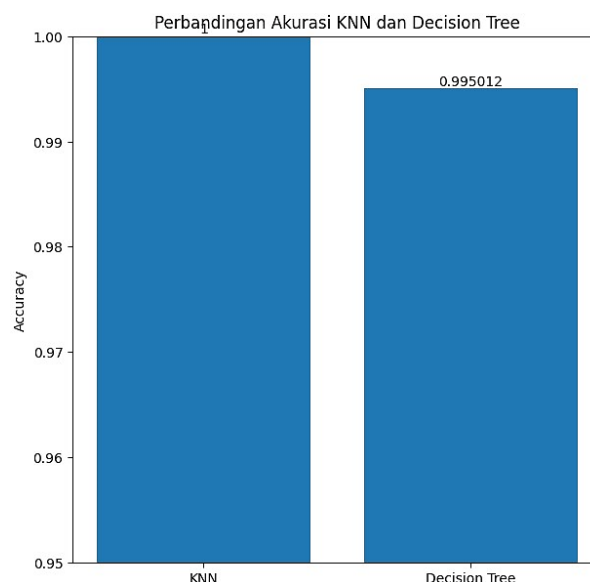
Tabel 5. Laporan klasifikasi model *Decision Tree*

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Low</i>	0.99	1.00	1.00	107
<i>Normal</i>	0.99	0.99	0.99	142
<i>High</i>	1.00	0.99	1.00	152
<i>Accuracy</i>			1.00	401
<i>Macro avg</i>	0.99	1.00	1.00	401
<i>Weighted avg</i>	1.00	1.00	1.00	401

3.4 Evaluation

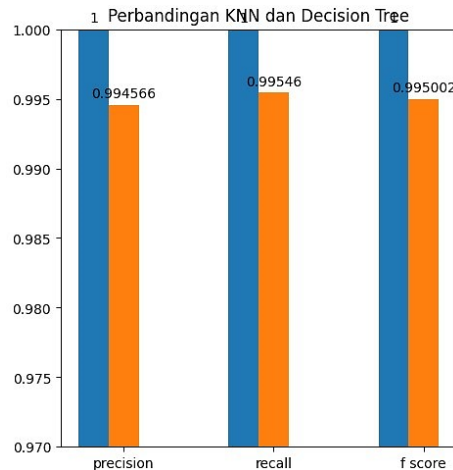
Pada tahap evaluasi ini, kinerja dari dua model klasifikasi, yaitu *K-Nearest Neighbor* (KNN) dan *Decision Tree*, dibandingkan menggunakan metrik evaluasi. Langkah pertama yang dilakukan adalah mengukur akurasi dari kedua model. Akurasi model KNN adalah 100%, sedangkan akurasi model *Decision Tree* adalah 99,50%. Hal ini menunjukkan bahwa kedua model mampu melakukan generalisasi yang baik terhadap data uji, dengan model KNN lebih unggul dalam hal akurasi.

Untuk visualisasi perbandingan akurasi kedua model, dibuat diagram batang. Diagram ini memperlihatkan bahwa akurasi model KNN dan *Decision Tree* sangat tinggi menandakan bahwa kedua model bekerja dengan sangat baik dalam klasifikasi tingkat *stress* individu.



Gambar 3. Diagram perbandingan KNN dan *Decision Tree*

Selanjutnya, metrik evaluasi lainnya seperti *precision*, *recall*, dan *f1-score* dihitung untuk kedua model. Untuk model KNN, *precision*, *recall*, dan *f1-score* adalah 100%. Sedangkan untuk model *Decision Tree*, *precision*, *recall*, dan *f1-score* masing-masing adalah 99,45%, 99,54%, dan 99,50%. Dari hasil ini dapat dilihat bahwa model KNN memiliki nilai *precision*, *recall*, dan *f1-score* yang paling tinggi, yang menunjukkan bahwa model KNN sangat efektif dalam melakukan prediksi.



Gambar 4. Diagram perbandingan akurasi KNN dan *Decision Tree*

Dalam hasil evaluasi Gambar 4, hasil *precision*, *recall*, dan *f1-score* kedua model dianalisis. Model KNN memiliki nilai *precision*, *recall*, dan *f1-score* sebesar 100%. Sedangkan model *Decision Tree* memiliki nilai *precision* sebesar 99,45%, *recall* sebesar 99,54%, dan *f1-score* sebesar 99,50%. Dari hasil tersebut, terlihat bahwa model KNN memiliki kinerja yang lebih baik dalam *precision*, *recall*, dan *f1-score* dibandingkan dengan model *Decision Tree*. Meskipun perbedaannya kecil, namun hal ini menunjukkan bahwa model KNN lebih baik dalam mengidentifikasi dan mengklasifikasikan tingkat *stress* dengan akurasi yang lebih tinggi dibandingkan dengan model *Decision Tree*.

Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model KNN lebih unggul dibandingkan dengan model *Decision Tree* dalam hal akurasi, *precision*, *recall*, dan *f1-score*. Namun, kedua model secara umum menunjukkan kinerja yang sangat baik dalam mengklasifikasikan tingkat *stress* individu. Keunggulan model KNN dalam *precision*, *recall* dan *f1-score* menunjukkan bahwa model ini lebih baik dalam memberikan prediksi yang benar positif dan memiliki keseimbangan yang baik antara *precision*, *recall*, dan *f1-score*.

4. KESIMPULAN

Berdasarkan hasil perbandingan antara algoritma KNN dan *Decision Tree* dalam melakukan klasifikasi tingkat *stress*, diperoleh bahwa algoritma KNN menunjukkan kinerja yang lebih baik dibandingkan dengan *Decision Tree*. Akurasi model KNN mencapai akurasi sempurna sebesar 100%, sedangkan *Decision Tree* mencapai akurasi sebesar 99,50%. Selain itu, nilai dalam *precision*, *recall*, dan *F1-score*, model KNN juga unggul dengan akurasi yang lebih tinggi daripada *Decision Tree*. Nilai akurasi *precision*, *recall*, dan *F1-score* masing-masing model KNN adalah 100%, sedangkan model *Decision Tree* memiliki nilai *precision*, *recall*, dan *F1-score* yang sedikit lebih rendah dengan masing-masing 99,45%, 99,54%, dan 99,50%. Selain itu, dari 2001 sampel yang dianalisis, mayoritas mengalami tingkat *stress* menengah hingga tinggi. Ditemukan bahwa suhu tubuh yang tinggi (diatas 33°C) berkorelasi dengan tingkat *stress* yang tinggi. Sementara kelembapan di atas 22,5 juga dikaitkan dengan *stress* tinggi. Pola unik terlihat pada langkah kaki, dimana jumlah langkah di bawah 90 dapat dikaitkan dengan tingkat *stress* normal dan tinggi, menunjukkan perlunya variabel lain untuk menentukan tingkat *stress* secara lebih akurat. Dari hasil penelitian, terbukti bahwa untuk tugas klasifikasi tingkat *stress*, algoritma KNN menjadi pilihan yang lebih tepat karena mampu memberikan prediksi yang lebih akurat dalam mengidentifikasi tingkat *stress* individu dan faktor suhu tubuh, kelembapan, dan jumlah langkah kaki memiliki peran signifikan dalam mempengaruhi tingkat *stress* individu. Oleh karena itu, penting bagi kita untuk lebih memperhatikan kondisi tubuh agar dapat mengelola *stress* dengan lebih baik di masa mendatang.

REFERENCES

- Ahmad, S. R., Anissa, M., Triana, R., & Kunci, K. (2022). HUBUNGAN TINGKAT STRES DENGAN KEJADIAN INSOMNIA PADA MAHASISWA ANGKATAN 2017 FAKULTAS KEDOKTERAN UNIVERSITAS BAITURRAHMAH. *Indonesian Journal for Health Sciences*, 6(1), 1–7.
- Alghifari, F., & Juardi, D. (2021). *Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes*.
- Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. In *IJCIT (Indonesian Journal on Computer and Information Technology)* (Vol. 6, Issue 2).
- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., Jajuli, M., HSRonggo Waluyo, J., Telukjambe Timur, K., Karawang, K., & Barat, J. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*. <https://doi.org/10.34010/jati.v12i1>



- Fardiana Risa, D., Pradana, F., & Abdurrachman Bachtiar, F. (2021). *IMPLEMENTASI METODE NAÏVE BAYES UNTUK MENDETEKSI STRES SISWA BERDASARKAN TWEET PADA SISTEM MONITORING STRES*. <https://doi.org/10.25126/jtiik.202184372>
- Fathirachman Mahing, N., Lazuardi Gunawan, A., Foresta Azhar Zen, A., Abdurrachman Bachtiar, F., Agung Wicaksono, S., Brawijaya, U., & Korespondensi, P. (n.d.). *KLASIFIKASI TINGKAT STRES DARI DATA BERBENTUK TEKS DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN RANDOM FOREST*. *10(7)*, 2023. <https://doi.org/10.25126/jtiik2023108010>
- Haganta Depari, D., Widiastiwi, Y., Mega Santoni, M., Ilmu Komputer, F., Pembangunan Nasional Veteran Jakarta, U., Fatmawati Raya, J. R., & Labu, P. (n.d.). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *JURNAL INFORMATIK Edisi Ke, 18*, 2022.
- Hambardzumyan, K. (2021). *Data Preprocessing in Real-time Education Management System*.
- Larose, D. T. (2005). *Discovering Knowledge In Data*. JohnWiley and Sons.
- Prabamurti, G. A. (2019). *Analisis Faktor-Faktor Pemicu Level Stress Akademik Mahasiswa Kedokteran Universitas Sebelas Maret Surakarta*.
- Purbolaksono, M. D., Irvan Tantowi, M., Imam Hidayat, A., & Adiwijaya, A. (2021). Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5(2)*, 393–399. <https://doi.org/10.29207/resti.v5i2.3008>
- Rachakonda, L., Mohanty, S. P., Koungianos, E., & Prabha Sundaravadivel. (2019). *Stress-Lysis: A DNN-Integrated Edge Device for Stress Level Detection in the IoMT*.
- Rahayu, P. W., & dkk. (2024). *Buku Ajar Data Mining* (Efitra, Ed.). Sonpedia.
- Suntoro, J. (2018). *Data Mining Algoritme dan Implementasi Menggunakan Bahasa Pemrograman PHP*.
- Tyas, T. M. M., & Purnamasari, A. I. (2023). Penerapan Algoritma K-means dalam Mengelompokkan Demam Berdarah Dengue Berdasarkan Kabupaten. *Blend Sains Jurnal Teknik*, *1(4)*, 277–283. <https://doi.org/10.56211/blendsains.v1i4.231>
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *4(2)*, 437. <https://doi.org/10.30865/mib.v4i2.2080>
- Van Fadhila, A., Azzahra, J. A., Rizki, K., Zulkarnain, T., Lathifah, N. D., Salsabiila, S. Z., & Chamidah, N. (2023). *Implementasi Metode Machine Learning Untuk Mendeteksi Tingkat Stres Manusia Berdasarkan Kualitas Tidur*.
- Wibowo, M., Rizieq, M., & Djafar, F. (2023). *JURNAL MEDIA INFORMATIKA BUDIDARMA Perbandingan Metode Klasifikasi Untuk Deteksi Stress Pada Mahasiswa di Perguruan Tinggi*. <https://doi.org/10.30865/mib.v7i1.5182>
- Wirayudha, V. R., Hidayat, N., & Dewi, R. K. (2020). *Identifikasi Tingkat Stress Pada Manusia Menggunakan Metode K-NN (K-Nearest Neighbour)* (Vol. 4, Issue 9). <http://j-ptiik.ub.ac.id>