

PENERAPAN MODEL ALGORITMA UNSUPERVISED LEARNING UNTUK KLASTERISASI TINGKAT KENYAMANAN RUANG TIDUR BERDASARKAN FAKTOR LINGKUNGAN

Mhd. Idham Khalif^{*1}, Listyo Edi Prabowo²

¹Universitas Trisakti, Jakarta Barat, ²Universitas Indonesia, Depok
Email: ¹idham.khalif@trisakti.ac.id, ²listyo.ediprabowo@gmail.com
^{*}Penulis Korespondensi

(Naskah masuk: 23 Januari 2024, diterima untuk diterbitkan: 12 April 2025)

Abstrak

Pada era revolusi industri 4.0, teknologi seperti *cloud computing*, *robotic*, *internet of things (IoT)*, *artificial intelligence (AI)* dan *machine learning (ML)* sangat banyak dikembangkan di berbagai sektor seperti industri, pemerintahan, pendidikan hingga ke sektor rumah tangga. Salah satu teknologi yang banyak digunakan di era revolusi industri 4.0 ini adalah AI atau ML. Pada penelitian ini telah berhasil menerapkan sebuah pengolahan data secara klusterisasi menggunakan model ML yaitu *unsupervised learning* menggunakan algoritma *k-means clustering* terhadap data yang sudah dikumpulkan. Data tersebut berupa data kuantitatif lingkungan ruang tidur yang dapat diukur, data tersebut diukur dan diambil menggunakan perangkat yang memiliki 5 sensor, yaitu sensor suara, cahaya, *temperature*, *humidity*, karbon dioksida (CO²), data nilai panjang, data nilai lebar serta, nilai cuaca yang diambil pada *platform OpenWheater* menggunakan perangkat yang dibangun berbasis IoT menggunakan koneksi internet dari sebuah *access point*. Data yang dikumpulkan menjadi sebuah dataset yang diolah menggunakan model algoritma *k-means clustering* sehingga dapat digunakan untuk melakukan klusterisasi terhadap kenyamanan ruang tidur. Dari hasil pengolahan menggunakan model algoritma *k-means clustering* terdapat nilai *Silhouette* sebesar 0,268 dengan persentase data per klaster adalah 29,92% untuk klaster 0 (tidak nyaman) dan 70,08% untuk klaster 1 (nyaman). Dalam praktik nyata, temuan dari hasil penerapan penelitian ini dapat digunakan untuk mengembangkan rekomendasi personalisasi untuk pengaturan ruang tidur, nantinya dapat merancang produk atau desain yang lebih efektif, di mana penerapan ML dapat dibenamkan secara langsung dalam sebuah sistem yang utuh. Dampaknya bisa sangat luas, mencakup peningkatan kesehatan, produktivitas, dan kesejahteraan penggunanya.

Kata kunci: *klusterisasi, unsupervised learning, kenyamanan, k-means clustering*

IMPLEMENTATION THE UNSUPERVISED LEARNING ALGORITHM MODEL FOR CLUSTERING LEVEL OF BEDROOM COMFORT LEVEL BASED ON ENVIRONMENTAL FACTORS

Abstract

In the era of Industry 4.0, technologies such as cloud computing, robotics, the Internet of Things (IoT), artificial intelligence (AI), and machine learning (ML) have been extensively developed across various sectors, including industry, government, education, and even households. One of the technologies widely used in this era is AI and ML. This research successfully applied clustered data processing using the ML model, specifically unsupervised learning with the *k-means clustering* algorithm, to the collected data. The data consists of quantitative measurements of the bedroom environment, gathered using a device with 5 sensors: sound, light, temperature, humidity, carbon dioxide (CO²), as well as length, width, and weather data taken from the OpenWeather platform using an IoT-based device with an internet connection from an access point. The collected data forms a dataset that is processed using the *k-means clustering* algorithm to classify the comfort level of the bedroom. The results of this processing showed a *Silhouette* value of 0.268, with the percentage of data per cluster being 29.92% for cluster 0 (uncomfortable) and 70.08% for cluster 1 (comfortable). In real practice, the findings from this research can be used to develop personalized recommendations for bedroom arrangements and to design more effective products. The application of ML can be embedded directly into a complete system, with far-reaching impacts, including improving users' health, productivity, and well-being.

Keywords: *clustering, unsupervised learning, comfort level, k-means clustering*

1. PENDAHULUAN

Kenyamanan adalah salah satu hal yang paling terpenting dalam kehidupan seorang individu manusia. Kenyamanan juga adalah salah satu faktor yang mempengaruhi kesehatan dan psikologi seorang individu. Karena kesehatan merupakan salah satu hal yang sangat penting bagi setiap individu, karena kesehatan adalah faktor utama yang memengaruhi semua aktivitas mereka (Khalif, Syauqy, & Maulana, 2018). Kenyamanan ruang tidur merupakan salah satu faktor yang mempengaruhi kualitas tidur seseorang. Pada dasarnya tidur adalah salah satu hal yang paling penting untuk menjaga kesehatan seseorang. Tidur memiliki peranan yang sangat penting bagi setiap individu, mulai dari anak-anak, remaja, dewasa, hingga lansia, karena dapat memengaruhi kualitas hidup secara keseluruhan, kesehatan fisik dan mental, serta cara berpikir, beraktivitas, belajar, dan berinteraksi dengan orang lain (Lisiswanti, Rodiani, Saputra, Sari, & Zafirah, 2019). Perubahan dalam aktivitas sehari-hari, tingkat kenyamanan lingkungan, serta kondisi mental atau pikiran individu, semuanya memiliki potensi besar untuk mempengaruhi kualitas tidur (Hutagalung, Marni, & Erianti, 2022). Banyak faktor yang saling berinteraksi dan dapat berdampak langsung pada kemampuan tubuh untuk beristirahat secara optimal. Elemen faktor tersebut memerlukan perhatian yang cermat, mengingat kualitas tidur yang buruk dapat berdampak jangka panjang pada kesehatan fisik dan mental seseorang.

Dalam beberapa jurnal penelitian menjelaskan ada beberapa faktor lingkungan yang mempengaruhi kenyamanan ruang tidur, seperti intensitas cahaya, suhu, tingkat kebisingan (Prabasari, Manungkalit, & Radjawane, 2022). Dari jurnal yang lain menyatakan faktor lingkungan berupa termal, seperti suhu dan kelembaban menjadi faktor yang mempengaruhi kualitas tidur (Cao, Lian, Ma, & Bao, 2021). Kualitas udara juga menjadi salah satu faktor yang mempengaruhi kenyamanan pada ruang tidur (Caddick, Gregory, Arsintescu, & Flynn-Evans, 2018). Berdasarkan jurnal penelitian yang sudah dilakukan dapat disimpulkan ada beberapa faktor lingkungan yang mempengaruhi kenyamanan ruang tidur, antara lain intensitas cahaya, tingkat kebisingan, suhu, kelembaban dan kualitas udara pada lingkungan ruang tidur.

Berdasarkan Standar Nasional Indonesia (SNI) nomor SNI-03-6575-2001 tentang tata cara perancangan sistem pencahayaan buatan pada bangunan gedung, tingkat cahaya yang ideal agar ruang tidur terasa nyaman berkisar 120 – 250 lux, lalu SNI-03-6386-2000 menerangkan bahwa tingkat bunyi yang ideal untuk ruang tidur adalah 30 – 35 db, SNI-03-6575-2001 menerangkan bahwa suhu ideal untuk ruang tidur agar terasa nyaman, berkisar 20,5°C – 27,1°C, untuk kelembaban ideal agar ruangan tidur terasa nyaman berkisar 40% - 60% dan untuk kualitas udara yang mengacu kepada *Federation of European Heating, Ventilation*

an Association (REHVA), berdasarkan kadar Karbondioksida (CO_2) agar ruang tidur terasa nyaman adalah berkisar 350 – 1000 ppm (Küçükhüseyin, & Özgür, 2021). Dapat disimpulkan bahwa untuk nilai tingkat kenyamanan ruang tidur berdasarkan faktor lingkungan.

Tabel 1. Nilai Tingkat Kenyaman Ruang Tidur Berdasarkan Faktor Lingkungan

Faktor Lingkungan	Batas Nilai Kenyamanan	Satuan
Intensitas Cahaya	100 - 250	lux
Tingkat Bunyi	30 - 35	db
Suhu	20,5 – 27,1	°C
Kelembaban	40 – 60	%
Kualitas Udara (CO_2)	350 - 1000	ppm
Kondisi Cuaca	<i>Clear Sky, Clouds, Rain, Thunderstorm, Haze</i>	-
Panjang	-	m
Lebar	-	m

Tabel 1 akan menjadi acuan sebagai nilai untuk melakukan klusterisasi pada model ML. Untuk melakukan klusterisasi pada data kenyamanan ruang tidur berdasarkan faktor lingkungan, salah satu teknik yang dapat digunakan adalah *unsupervised learning*. Teknik ini memungkinkan pengelompokan data berdasarkan kesamaan karakteristik tanpa memerlukan label atau kategori yang telah ditentukan sebelumnya. Dalam konteks ini, metode *clustering* yang digunakan dalam *unsupervised learning* dapat membantu mengidentifikasi pola atau kelompok kenyamanan yang berbeda berdasarkan faktor-faktor lingkungan yang terukur. Metode *unsupervised learning* adalah salah satu metode algoritma yang dapat mengelompokkan pola serupa berdasarkan fitur data (Song, Hwang, Kim, Kim, Lee, Park, & Cho, 2024). Banyak sekali metode algoritma pada *unsupervised learning* diantaranya *k – means clustering* dan *hierarchical clustering*.

Pada penelitian ini bertujuan untuk melakukan penerapan terhadap model *unsupervised learning* berupa *k-means clustering* yang dapat digunakan untuk klusterisasi kenyamanan ruang tidur berdasarkan faktor lingkungan. Metode tersebut akan dilihat dan diuji untuk mengevaluasi efektivitas dalam membentuk kelompok - kelompok tingkat kenyamanan yang relevan. Melalui analisis ini, diharapkan dapat diperoleh model yang optimal untuk memahami pengaruh faktor-faktor lingkungan terhadap kenyamanan ruang tidur, yang pada nantinya dapat memberikan rekomendasi bagi perancangan sebuah sistem utuh yang ditanami sebuah model ML untuk melakukan monitoring kenyamanan ruang tidur secara *real-time*.

2. PENELITIAN TERKAIT

Dalam penelitian yang pernah dilakukan terhadap analisa hubungan nilai kondisi lingkungan pada kamar tidur, seperti PM2.5, CO2, suhu, kelembaban, dan kebisingan terhadap kualitas kenyamanan dan kualitas tidur. Telah mampu membuat rekomendasi dalam kenyamanan dan kualitas tidur, data yang diambil juga telah menggunakan sebuah alat yang dapat mengukur nilai tersebut, namun dalam penentuan rekomendasi, masih secara subjektif berdasarkan survei harian yang rentan terhadap bias persepsi peserta, yang dapat memengaruhi kesimpulan tentang kenyamanan dan kualitas tidur (Basner, Jones, Ecker, Howard, Schneller, Cordoza, Kaizi-Lutu, Park-Chavar, Stahn, Dinges, Shou, Mitchell, Bhatnagar, Smith, Smith, Stopforth, Yeager, & Keith, 2023).

Pada penelitian yang pernah dilakukan oleh (Suryadi Tatang, & Sri, 2020) mampu membuat sebuah sistem cerdas untuk memantau kenyamanan ruang kelas berbasis *Internet of Things* (IoT) yang ditanami beberapa sensor. Pada penelitian ini menggunakan model algoritma *k – means clustering* dengan fitur nilai kelembaban ruangan, suhu, suara dan cahaya. Hasil yang didapat model algoritma *k-means clustering* mampu melakukan klusterisasi dengan baik. Namun pada penelitian ini jumlah dataset yang digunakan terbilang cukup sedikit sebanyak 20 baris dataset.

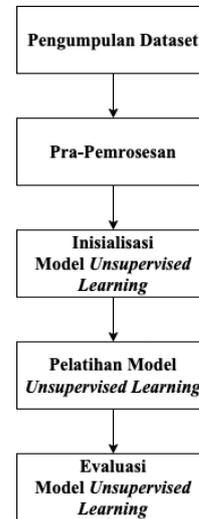
Penelitian yang kedua dilakukan oleh (Li, Liu, Chen, & Xue, 2024). Penelitian ini melakukan efisiensi terhadap data termal kenyamanan ruangan *indoor* menggunakan metode *transfer learning*. Penelitian yang dilakukan metode yang dilakukan mudah pengimplementasiannya dan hasilnya dengan metode yang digunakan kinerja untuk klusterisasi yang dilakukan lebih baik dibandingkan dengan metode yang tradisional.

Penelitian ketiga dilakukan oleh (Liu, Li, Sun, Dong, Yin & Yan, 2024) di mana pada penelitian ini melakukan prediksi terhadap kenyamanan termal di dalam ruangan menggunakan model algoritma *k-means clustering*. Pada penelitian ini akurasi dari model prediksi *k-means clustering* mencapai 90 % dengan 3 atribut utama yaitu *clothing (CLO)*, *Air Temperature (TA)*, *Relative Humidity (RH)*.

Dari uraian beberapa penelitian yang sudah ada, pada penelitian ini akan melakukan pengolahan data lingkungan ruang tidur menggunakan model *unsupervised learning* yaitu *k-means clustering* untuk melakukan klusterisasi kenyamanan ruang tidur. Berdasarkan review dari penelitian terkait belum ada penerapan model *unsupervised learning* yaitu *k-means clustering* untuk melakukan klusterisasi kenyamanan ruang tidur tersebut, tentunya hasil dari klusterisasi tersebut lebih objektif dibandingkan dengan survei yang dilakukan.

3. METODE PENELITIAN

Pada penelitian ini ada beberapa tahapan penelitian yang dilakukan, terdiri dari, pengumpulan dataset, pra-pemrosesan data, inialisasi model *unsupervised learning*, pelatihan model *unsupervised learning*, dan evaluasi model *unsupervised learning*.



Gambar 1. Alur tahapan penelitian

3.1. Pengumpulan Dataset

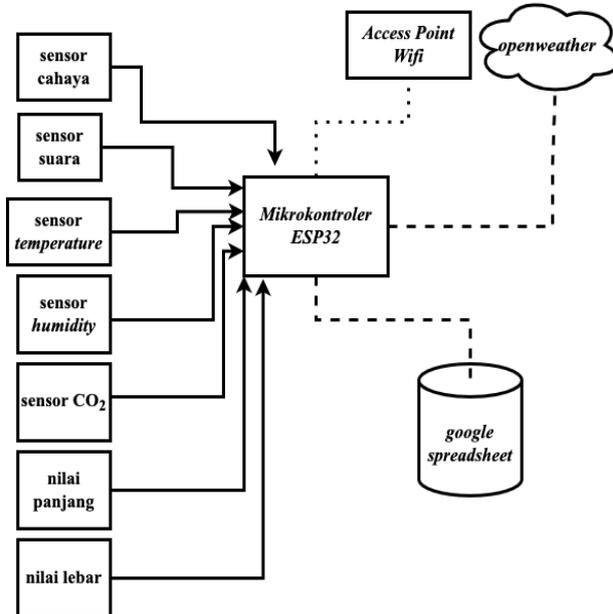
Teknik pengumpulan data merujuk pada metode yang digunakan oleh peneliti untuk memperoleh informasi atau data yang dibutuhkan dalam penelitian dan merupakan langkah yang sangat penting dalam metodologi penelitian (Daruhadi, & Pia, 2024) Dataset memiliki semua karakteristik, fitur, dan fungsi yang dimiliki oleh basis data pada umumnya. Dataset dapat terdiri dari beberapa tabel, dan tabel-tabel tersebut dapat saling terhubung (*relationship*) (Yahya, & Mahpuz, 2019).

Pada penelitian ini dataset adalah komponen yang paling penting dan harus ada, karena dataset adalah faktor utama mode *unsupervised learning* dapat bekerja. Penelitian ini menggabungkan data kuantitatif dan kualitatif, yang di mana data kuantitatif berupa nilai suhu, kelembaban, tingkat kebisingan, intensitas cahaya, kualitas udara (kadar CO₂), panjang dan lebar. Untuk data kualitatif berupa data cuaca. Pengumpulan dataset dilakukan dengan pengurukan langsung dengan sebuah perangkat yang memiliki sensor dan menghasilkan kumpulan data yang menjadi dataset.

1). Pengukuran data

Dataset yang digunakan untuk diolah, diambil menggunakan satu unit perangkat pengukur berbasis IoT. Arsitektur proses pengukuran ini akan menggambarkan bagaimana data di-*sensing*, data dikirim dan disimpan pada *google spreadsheet*. Untuk mengambil data dari *openwheater* dan menyimpan data ke *google spreadsheet* membutuhkan sebuah koneksi internet menggunakan

sebuah *access point*. Gambar 2 adalah desain arsitektur yang diterapkan pada proses pengukuran yang dilakukan untuk memperoleh dataset yang digunakan dan nantinya diolah oleh model *unsupervised learning*.



Gambar 2. Arsitektur proses pengukuran data

Gambar 2 adalah arsitektur perangkat pengambilan data sensor dan cuaca pada *openweather*, serta pengiriman data pada *Google Spreadsheet* untuk disimpan. Pada perangkat pengukur terdapat 6 sensor yang terhubung, di mana perangkat tersebut terkoneksi menggunakan jaringan internet *hotspot Wi-Fi* untuk mengambil data dari *openweather* dan mengirim data ke *Google Spreadsheet*. Setelah semua nilai data terkumpul, data yang sudah didapat akan dikirim menggunakan koneksi internet ke *Google Spreadsheet*, dataset yang sudah didapat bisa dikonversi ke file *csv/xlsx* atau dibaca langsung untuk diolah dengan model *unsupervised learning*.

Pada perangkat pengukur, terdapat 8 data yang diambil di mana, memiliki 5 sensor, nilai Panjang & lebar, serta nilai yang diambil dari *platform openweather*. Sensor cahaya menggunakan BH1750, sensor suhu dan kelembaban menggunakan DHT11, sensor kualitas udara menggunakan MQ-135 dan nilai keadaan cuaca yang diambil dari data *openweather* secara *real-time*. Perangkat pengukur terdapat mikrokontroler ESP32 sebagai media untuk memproses, membaca nilai sensor, dan mengirimkan secara *real-time* ke *Google Spreadsheet* untuk disimpan, dapat dilihat pada Gambar 3.

Untuk nilai panjang dan lebar ruangan sudah ditentukan secara langsung. Pada penelitian ini terdapat 3 jenis ukuran ruang tidur, yaitu ukuran dengan Panjang 3 meter dan lebar 3 meter, panjang 3 meter dan lebar 4 meter, serta ukuran panjang 4 meter dan lebar 4 meter. Sehingga pada penelitian ini,

dataset yang digunakan terdapat 8 atribut, yaitu intensitas cahaya dalam satuan *lux (lx)*, tingkat kebisingan (suara) dalam satuan *decibel (db)*, suhu dalam satuan derajat selsius ($^{\circ}\text{C}$), kelembaban dalam satuan persen (%), kualitas udara (CO_2) dalam satuan *part per million (ppm)*, cuaca, panjang dan lebar dalam satuan meter (m). Dengan menggunakan perangkat yang dirancang ini akan menghasilkan kumpulan data yang menjadi dataset untuk pengolahan model.



Gambar 3. Perangkat pengukur nilai kondisi lingkungan ruang tidur

3.2. Pra-Pemrosesan Data

Pra-pemrosesan data adalah salah satu langkah yang dilakukan untuk mempersiapkan dataset dapat digunakan secara efektif dalam model algoritma yang digunakan. Dalam penerapan ML atau proses pengolahan data lainnya tahapan pra-pemrosesan data sangat penting dilakukan.

Pada penelitian ini tahapan pra-pemrosesan data dilakukan untuk mempersiapkan dataset yang digunakan agar dapat diolah dengan baik dan mendapatkan hasil pengolahan yang optimal. Berikut ini ada beberapa tahapan pra-pemrosesan data yang dilakukan terhadap data yang sudah dikumpulkan.

1). *Encoding* kategorikal data

Encoding kategorikal data adalah suatu tahapan merubah satu kolom atau beberapa kolom dataset yang bertipe data objek menjadi angka. Tujuan tahapan ini dilakukan karena pada pengolahan ML menggunakan *k-means clustering* hanya dapat mengolah dataset dalam bentuk angka saja.

Pada penelitian ini, dataset yang telah dikumpulkan terdapat atribut berupa data kategorikal yaitu cuaca, di mana data cuaca memiliki jenis data berupa karakter huruf. Agar data cuaca dapat diolah oleh model *unsupervised learning* menggunakan

algoritma *k-means clustering*, kita perlu melakukan *encoding* terhadap data tersebut, yaitu merubah data cuaca menjadi pengkodean dalam bentuk nilai angka. Teknik yang digunakan adalah *label encoding* dengan memberikan label nilai angka untuk setiap kategori dalam satu kolom. Dapat dilihat pada Tabel 2.

Tabel 2. *Label encoding* pada atribut cuaca

Label	Nilai <i>encoding</i>
<i>Clear Sky</i>	1
<i>Clouds</i>	2
<i>Rain</i>	3
<i>Thunderstorm</i>	4
<i>Haze</i>	5

2). Penanganan data duplikasi (*data cleansing*)

Pada proses penanganan data duplikasi, (*data cleansing*) tujuan utamanya adalah untuk menghilangkan data *entry* yang ter-duplikat sehingga dapat mempengaruhi kualitas dataset secara keseluruhan. *Data cleansing* adalah salah satu teknik dalam data mining yang bertujuan untuk menghasilkan basis data yang berkualitas dan bebas dari duplikasi (Rokhman, Ningtyas, Salim, & Santoso, 2020).

Data yang duplikat bisa muncul akibat kesalahan pengambilan data atau pengulangan pencatatan yang tidak diperlukan dan jika dibiarkan, hal ini dapat mengarah pada menurunkan akurasi model ML yang akan diterapkan. Oleh karena itu, dalam tahap ini, langkah yang dilakukan adalah mendeteksi dan memeriksa data yang serupa atau identik, baik di seluruh baris atau hanya pada kolom tertentu yang relevan. Selanjutnya, data duplikat tersebut akan dihapus, sehingga hanya tersisa satu entri untuk setiap kombinasi nilai yang unik. Proses ini biasanya akan mengurangi jumlah data yang tersedia, yang dikenal dengan istilah penyusutan dataset, namun secara keseluruhan meningkatkan kualitas data yang lebih representatif dan dapat diandalkan.

Dalam penelitian ini tahapan penanganan data duplikasi perlu dilakukan dikarenakan proses pengambilan dan pengumpulan data menggunakan perangkat pengukur berbasis IoT. Perangkat IoT umumnya mengumpulkan data secara *periodic*, dalam penelitian ini, data yang diambil pada interval waktu 10 menit, mungkin tidak mengalami perubahan signifikan dari satu periode pengukuran ke periode berikutnya. Hal ini menyebabkan sensor atau perangkat pengukur, mencatat nilai yang sama berulang kali, sehingga menyebabkan adanya data yang tidak berubah atau nilai yang tetap yang dihasilkan oleh sensor dalam periode tersebut. Oleh karena itu, dalam penelitian ini, untuk tahapan skenario pengumpulan data, penanganan duplikasi sangat krusial dan harus dilakukan untuk memastikan bahwa dataset yang digunakan tidak hanya akurat,

tetapi juga tidak terdistorsi oleh pembacaan yang berulang atau tidak relevan.

3). Pengacakan data (*data shuffling*)

Pengacakan data (*data shuffling*) adalah suatu proses mengacak urutan data dalam dataset secara keseluruhan untuk memastikan bahwa data tidak terurut atau terstruktur dalam satu pola yang sama sehingga dapat mempengaruhi hasil analisis atau model algoritma yang diterapkan. Tujuan dari dilakukannya pengacakan data (*data shuffling*) adalah untuk mengurangi bias yang terjadi terhadap pola data yang sama secara berturut-turut.

Pada penelitian ini, tahapan pengacakan data (*data shuffling*) perlu dilakukan karena data yang diambil menggunakan perangkat pengukur dengan beberapa sensor yang memiliki waktu jarak pengambilan data, yaitu 10 menit secara terus-menerus, sangat memungkinkan terjadinya pola-pola data tertentu yang berurutan. Data yang tidak terstruktur pada pengolahan ML dapat mempermudah model dalam mengenali pola yang kompleks pada dataset. Dengan demikian, penelitian ini menerapkan pengacakan data (*data shuffling*) agar data yang disajikan lebih baik dan mengoptimalkan kinerja dari model yang diterapkan.

3.3. Inisialisasi Model *Unsupervised Learning*

Pada penelitian ini menerapkan satu model ML, *unsupervised learning* yaitu algoritma *k-means clustering* yang digunakan untuk pengolahan dataset. *K-means clustering* adalah salah satu metode algoritma yang digunakan untuk mengelompokkan data ke dalam beberapa bagian, di mana metode ini berbasis jarak dan membagi data menjadi beberapa kluster. *K-means clustering* hanya dapat diterapkan pada atribut numerik dan termasuk dalam kategori *partitioning clustering*, yang memisahkan data ke dalam k kelompok yang berbeda (Febrianto, Achmadi, & Sasmito, 2021). Gambar 4 adalah *flowchart* langkah – langkah penerapan model algoritma *k-means clustering*.

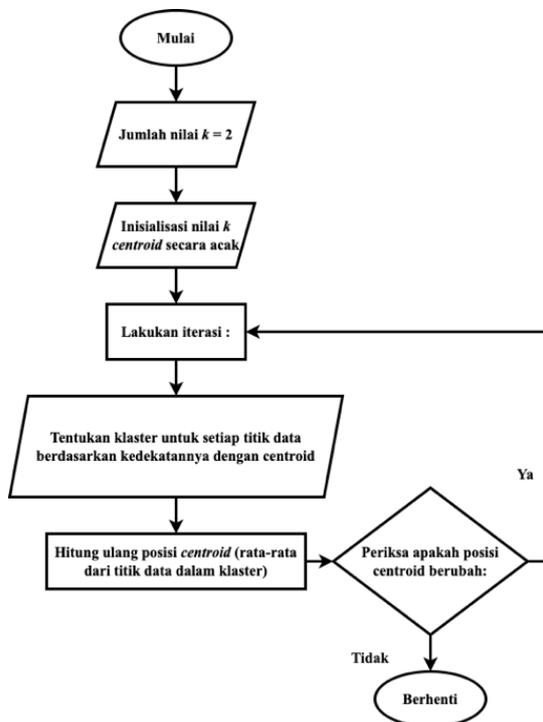
1) Inisialisasi nilai kluster (*centroid*)

Pada langkah pertama ini adalah melakukan inisialisasi terhadap kluster (*centroid*) yang digunakan dan menentukan nilai *k*, sehingga algoritma *k-means clustering* akan menginisialisasi nilai *k* pusat kluster secara acak. Setiap *centroid* adalah titik di ruang fitur yang mewakili pusat dari kluster tersebut. Pada tahapan ini ada 2 kluster (*centroid*) atau nilai *k* = 2, yang digunakan yaitu kluster “nyaman” dan klusterisasi “tidak nyaman”.

2). Iterasi proses

Iterasi proses adalah suatu proses pada model algoritma *k-means clustering*, yang mengulangi langkah atau proses tertentu. Pada algoritma *k-means clustering* yang diterapkan pada penelitian ini, proses

yang dilakukan *iterasi* adalah proses pada poin 3 dan 4 berikut ini. Proses ini akan terus berulang sampai memenuhi kondisi tidak ada perubahan dan konvergensi tercapai.



Gambar 4. Flowchart tahapan penerapan algoritma model *k-means clustering*

3) Menentukan titik data kluster

Tahapan ini bertujuan untuk menghitung jarak setiap kluster (*centroid*) menggunakan persamaan *Euclidean* dan nantinya akan mendapatkan nilai jarak dengan kluster (*centroid*) tertentu, Persamaan 1 adalah persamaan *Euclidean* yang digunakan untuk menghitung antara dua titik $A(x_1, y_1)$ sampai $B(x_8, y_8)$ karena pada penelitian ini memiliki 8 atribut dan nantinya akan mendapatkan nilai jarak antara A dan B, untuk menjadi patokan pengklusterisasian. Nilai x dan y merujuk pada koordinat atribut dari dua titik data yang berbeda. Lebih spesifik lagi, dalam kasus ini, titik A dan titik B merupakan dua titik data yang akan dihitung jaraknya satu sama lain.

$$d(A, B) = \sqrt{(x_1 - y_1)^2 + \dots + (x_8 + y_8)^2} \quad (1)$$

Persamaan 1 memiliki nilai x dan y , di mana x adalah nilai atribut vektor pertama, sedangkan nilai y adalah nilai atribut vektor kedua yang ingin dibandingkan. Pada persamaan 1 terdapat notasi persamaan untuk mewakili setiap 8 atribut yang dimiliki pada dataset yaitu $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5), (x_6, y_6), (x_7, y_7), (x_8, y_8)$ di mana, notasi (x_1, y_1) adalah nilai dari intensitas cahaya, (x_2, y_2) adalah nilai dari tingkat bunyi, (x_3, y_3) adalah nilai dari suhu, (x_4, y_4) adalah nilai dari kelembaban, (x_5, y_5) adalah nilai dari kualitas

udara, (x_6, y_6) adalah nilai dari kondisi cuaca, (x_7, y_7) adalah nilai dari panjang ruang tidur dan (x_8, y_8) adalah nilai dari lebar ruang tidur.

4). Menghitung ulang nilai kluster (*centroid*)

Pada tahap ini akan dilakukan perhitungan ulang terhadap nilai kluster (*centroid*) tujuannya adalah memperbaiki dan memperbarui posisi pusat setiap kluster setelah titik data dikelompokkan berdasarkan kedekatannya dengan *centroid* pada iterasi sebelumnya. Dengan kata lain, menghitung ulang *centroid* membantu algoritma untuk menyempurnakan pembagian kluster sehingga lebih akurat dan lebih sesuai dengan distribusi data. Di mana kluster (*centroid*) baru untuk kluster tersebut akan menjadi rata-rata dari nilai-nilai tersebut untuk setiap fitur, menggunakan persamaan 2 berikut ini.

$$C_k = \frac{1}{n_k} \sum d_1 d_2 \dots d_8 \quad (2)$$

Pada Persamaan 2 di atas nilai C_k adalah *centroid* kluster ke- k . Nilai k adalah 2 yang mewakili kluster 1 dan kluster 2. Notasi n_k adalah jumlah data dalam kluster ke- k . Notasi persamaan d_1 hingga d_8 adalah nilai titik data rekam dalam jarak setiap fitur yang telah dihitung sebelumnya. Algoritma *k-means clustering* akan melakukan pengecekan terhadap perubahan nilai kluster (*centroid*).

5). Algoritma berhenti

Pada tahapan pemrosesan model algoritma *k-means clustering* yang digunakan, pada penelitian ini akan berhenti setelah proses konvergensi yaitu ketika tidak ada perubahan lebih lanjut dalam pembagian data ke kluster atau perubahan posisi *centroid* dan nantinya hasil dari klusterisasi menggunakan model algoritma *k-means clustering* dapat divisualkan dalam bentuk grafik untuk melihat sebaran data yang sudah diklusterisasikan.

4. HASIL DAN PEMBAHASAN

4.1. Dataset

Dataset yang digunakan pada penelitian ini adalah dataset diambil secara langsung dan *realtime* menggunakan perangkat pengukur nilai kondisi ruang tidur yang dapat dilihat pada Gambar 3 di atas, lalu dataset tersebut dilakukan pra-pemrosesan data, seperti *encoding* atribut yang berupa data karakter, penanganan data duplikasi (*data cleansing*), dataset yang didapat sebanyak 1010 baris lalu dilakukan penanganan data duplikasi (*data cleansing*) sehingga dataset menyusut menjadi 822 baris, setelah data duplikasi ditangani, selanjutnya melakukan pengacakan (*data shuffling*) terhadap dataset agar data yang disajikan lebih baik dan tidak terpola.

Tabel 3. Sampel dataset yang digunakan

No	lux	db	°C	%	ppm	cuaca	m	m
1	74	74	30,8	75	816	1	3	4
2	75	58	30,8	75	786	2	3	4
3	100	38	26,5	50	757	3	3	4
...
826	101	36	26,5	76	708	5	3	3
827	25	58	30,8	82	934	5	3	3
828	95	77	29,8	76	672	3	4	4

Tabel 3 adalah sampel dari dataset yang mewakili setiap jenis bentuk data dari setiap nilai sensor yang sudah dilakukan proses pra-pemrosesan. Terdapat 8 kolom utama yang akan diolah dengan model algoritma *k-means clustering*. Kolom ke-1 adalah *lux* berupa nilai dari intensitas cahaya, kolom ke-2 adalah *db* merupakan nilai dari suara, kolom ke-3 adalah °C nilai dari suhu, kolom ke-4 adalah % merupakan nilai dari kelembaban, kolom ke-5 adalah ppm merupakan nilai kadar karbon dioksida (CO₂), kolom ke-6 adalah cuaca yang telah dilakukan *encoding*, kolom ke-7 adalah m merupakan nilai panjang dan kolom ke-8 adalah m merupakan nilai dari lebar. Semua data pada Tabel 3 di atas adalah hasil konkret dari proses pra-pemrosesan data yang dilakukan untuk pemrosesan data menggunakan *k-means clustering* menjadi lebih baik.

4.2. Hasil Penerapan Klasterisasi K-Means Clustering

Bagian ini adalah bagian pembahasan hasil penerapan model ML yaitu *k-means clustering* secara teknis menggunakan bahasa pemrograman. Dari penerapan model ML yang dilakukan, bahasa pemrograman yang digunakan adalah bahasa pemrograman *python* untuk penerapan model ML secara keseluruhan, untuk *integrated development environment (IDE)* menggunakan *Google Collab (Google Collaboratory)* yang dianggap lebih mudah digunakan karena tidak perlu melakukan *installasi* aplikasi ke dalam komputer pribadi, karena sudah berbasis halaman *web* dan juga mendukung bahasa pemrograman *python*, dan dataset yang digunakan, dalam format *.csv* yang dapat di-*import* ke dalam kode program *python* yang dapat diolah, sehingga memiliki beberapa hasil yang didapat.

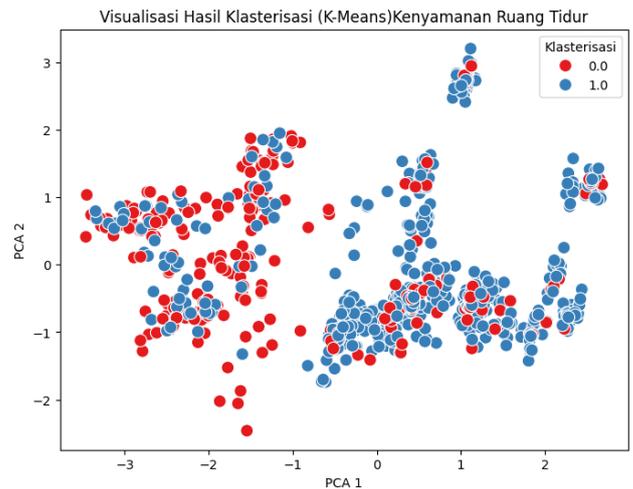
Gambar 5 adalah hasil visual klasterisasi yang dihasilkan oleh pengolahan model algoritma *k-means clustering*. Dari Gambar 5 model algoritma *k-means clustering* sudah mampu melakukan klasterisasi atau mengelompokkan terhadap dataset lingkungan kamar tidur, dengan iterasi sebanyak 10 kali, di mana pada visual berwarna merah adalah klaster 0 yaitu tidak nyaman, sedangkan untuk visual berwarna biru

adalah klaster 1 yaitu nyaman, jika hasil klasterisasi dibuat dalam bentuk tabel.

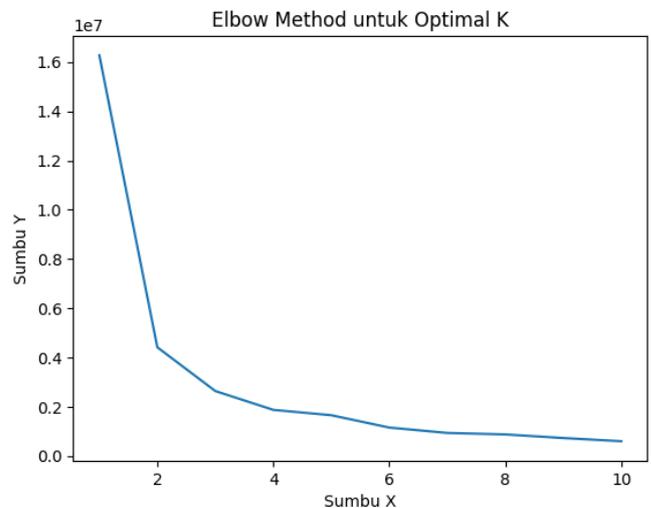
Tabel 4. Hasil klasterisasi model algoritma *k-means clustering*

No	lux	db	°C	%	ppm	cuaca	m	m	C
1	74	74	30,8	75	816	1	3	4	0
2	75	58	30,8	75	786	2	3	4	0
3	100	38	26,5	50	757	3	3	4	1
...
826	101	36	30,8	78	708	3	4	4	1
827	25	58	30,8	82	924	5	3	3	0
828	95	77	29,8	76	672	3	4	4	1

Tabel 4 di atas, adalah hasil pengolahan model algoritma *k-means clustering* dalam bentuk Tabel. Pada kolom terakhir, terdapat kolom C yang merupakan klasterisasi yang didapat. Jika merujuk Tabel 1 di atas, didapat hasil yang sesuai untuk



Gambar 5. Visualisasi hasil klasterisasi *K-Means clustering*



Gambar 6. Visualisasi hasil nilai *elbow method*

klasterisasi 0 untuk tidak nyaman dan 1 untuk nyaman.

Hasil lainya yang didapat adalah sebaran persentase data per klaster, di mana untuk klaster 0 (tidak nyaman) persentasenya sebesar 29,93% untuk klaster 0 (Tidak nyaman) dan untuk klaster 1 (nyaman) persentase datanya adalah 70,07%. Lalu didapat juga *silhouette score*. Metode evaluasi internal seperti *silhouette score* telah menjadi fokus penelitian dalam evaluasi *clustering* (Suraya, Soleh, & Lestari, 2023). Tujuan dari *silhouette score* atau yang disebut dengan *silhouette coefficient* merupakan sebuah metode untuk melakukan pengukuran terhadap kualitas dan kekuatan klasterisasi (Handoyo, Mangkudjaja, & Michrandi Nasution, 2014). Di mana nilai *silhouette score* kriteria sangat baik adalah $\geq 0,7$, cukup baik pada rentang nilai 0,2–0,7. Pada hasil pemrosesan menggunakan model algoritma *k-means clustering* ini nilai *silhouette* adalah 0,268, yang di mana cukup baik namun bisa dioptimalkan kembali.

Gambar 6 adalah visualisasi grafik *elbow method* yang digunakan untuk menentukan jumlah klaster optimal dalam penerapan model *k-means clustering*. Dapat dilihat sumbu X adalah menunjukkan jumlah klaster dan sumbu Y adalah menunjukkan nilai *inertia* atau *within-cluster sum of squares* (WCSS). Semakin kecil nilai *inertia*, semakin baik pembagian klaster. Visualisasi hasil *elbow method* untuk optimal nilai *k* pada Gambar 6 memperlihatkan hasil yang cukup baik, dikarenakan nilai dari WCSS dari waktu ke waktu *iterasi* semakin mengecil.

5. KESIMPULAN

Penelitian ini berhasil mengambil dan mengumpulkan dataset pada ruang tidur, secara langsung menggunakan perangkat pengukur dengan jarak waktu *sampling* adalah 5 menit, dengan jumlah 1010 baris dan menyusut menjadi 822 baris, karena proses penanganan duplikasi (*data cleansing*), lalu dataset tersebut sudah melalui proses pra-pemrosesan, lalu diolah menggunakan model algoritma *k-means clustering*. Model algoritma *k-means clustering* yang digunakan untuk melakukan klasterisasi terhadap dataset lingkungan ruang tidur telah mampu melakukan klasterisasi atau pengelompokan dengan 2 klaster yaitu klaster 0 untuk nilai tidak nyaman dan klaster 1 untuk nilai nyaman. Nilai *silhouette* yang didapat adalah sebesar 0,268. Sebaran persentasenya adalah sebesar 29,93% untuk klaster 0 (Tidak nyaman) dan untuk klaster 1 (nyaman) persentase datanya adalah 70,07%. Kedepanya penelitian ini bisa menjadi sebuah rekomendasi untuk penerapan sebuah sistem yang utuh, untuk melakukan monitoring secara *real-time* kenyamanan ruang tidur menggunakan ML dan ditambahkan aksi yang dibutuhkan untuk meningkatkan kenyamanan ruang tidur.

DAFTAR PUSTAKA

- BASNER, M., SMITH, M. G., JONES, C. W., ECKER, A. J., HOWARD, K., SCHNELLER, V., CORDOZA, M., KAIZI-LUTU, M., PARK-CHAVAR, S., STAHN, A. C., DINGES, D. F., SHOU, H., MITCHELL, J. A., BHATNAGAR, A., SMITH, T., SMITH, A. E., STOPFORTH, C. K., YEAGER, R., & KEITH, R. J., 2023. Associations of bedroom PM2.5, CO2, temperature, humidity, and noise with sleep: An observational actigraphy study. *Sleep Health*, 9(3), 253–263. <https://doi.org/https://doi.org/10.1016/j.sleh.2023.02.010>
- CADDICK, Z. A., GREGORY, K., ARSINTESCU, L., & FLYNN-EVANS, E. E., 2018. A review of the environmental parameters necessary for an optimal sleep environment. *Building and Environment*, 132, 11–20. <https://doi.org/10.1016/j.buildenv.2018.01.020>
- CAO, T., LIAN, Z., MA, S., & BAO, J., 2021. Thermal comfort and sleep quality under temperature, relative humidity and illuminance in sleep environment. *Journal of Building Engineering*, 43, 102575. <https://doi.org/10.1016/j.job.2021.102575>
- DARUHADI, GAGAH., SOPIATI, PIA., 2024. Pengumpulan Data Penelitian. -CEKI : Jurnal Cendekia Ilmiah Vol.3, No.5. pp. 5423-5443.
- FEBRIANTO, A., ACHMADI, S., & SASMITO, A. P., 2021. Penerapan Metode K-Means untuk Clustering Pengunjung Perpustakaan Itn Malang. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 5, Issue 1). <https://doi.org/10.36040/jati.v5i1.3222>
- HANDOYO, R., RUMAMI M, R., & NASUTION, S. M., 2014. Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K - Means Pada Pengelompokan Dokumen. *JSM STMIK Mikroskil*, 73- 82.
- HUTAGALUNG, N., MARNI, E. . AND ERIANTI, S., 2022. Faktor-Faktor Yang Mempengaruhi Kualitas Tidur Pada Mahasiswa Tingkat Satu Program Studi Keperawatan Stikes Hang Tuah Pekanbaru. *Jurnal Keperawatan Hang Tuah (Hang Tuah Nursing Journal)*, 2(1), pp. 77–89. <https://doi.org/10.25311/jkh.Vol2.Iss1.535>
- KHALIF, M. I., SYAUQY, D., & MAULANA, R., 2017. Pengembangan Sistem Penghitung Langkah Kaki Hemat Daya Berbasis Wemos D1 Mini. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(6), 2211–2220. Diambil dari

- ptiik.ub.ac.id/index.php/j-ptiik/article/view/1567
- KÜÇÜKHÜSEYİN, ÖZGÜR. 2021. CO₂ Monitoring And Indoor Air Quality, The REHVA European HVAC Journal, pp. 54 – 59.
- LI, K., LIU, Y., CHEN, L., & XUE, W., 2024. Data efficient indoor thermal comfort prediction using instance based transfer learning method. *Energy and Buildings*, 306, 113920. <https://doi.org/10.1016/j.enbuild.2024.113920>
- LIU, Y., LI, X., SUN, C., DONG, Q., YIN, Q., & YAN, B., 2024. An indoor thermal comfort model for group thermal comfort prediction based on K-means++ algorithm. *Energy and Buildings*, 115000. <https://doi.org/10.1016/j.enbuild.2024.115000>
- LISISWANTI, R., SAPUTRA, O., INDAH SARI, M., & HANA ZAFIRAH, N., 2019. Hubungan Antara Kualitas Tidur terhadap Hasil Belajar Mahasiswa Program Studi Pendidikan Dokter Fakultas Kedokteran Universitas Lampung. In *J Agromedicine Volume (Vol. 6)*.
- PRABASARI, NINDA AYU., MANUNGKALIT, MARIA. & RADJAWANE, CORMUTIA JESICA., 2022. Lingkungan di Panti Werdha yang Mendukung Kualitas Tidur Baik pada Lanjut Usia (Environment in nursing Home That Supports Good Sleep Quality In The Elderly). *Lingkungan di Panti Werdha yang Mendukung Kualitas Tidur Baik pada Lanjut Usia (Environment in nursing Home That Supports Good Sleep Quality In The Elderly)*, 10 (2). pp. 60-67. ISSN pISSN: 2338-624X, eISSN: 2614-4352.
- ROKHMAN, N., NINGTYAS, A. M., SALIM, M. F., & SANTOSO, D. B., 2020. Penerapan Sistem Data Cleansing untuk Mencegah dan Menghilangkan Duplikasi Rekam Medis. *Jurnal Pengabdian Kepada Masyarakat (Indonesian Journal of Community Engagement)*, 6(4). <https://doi.org/10.22146/jpkm.51073>
- SNI-03-6575-2001. Tata Cara Sistem Pencahayaan Buatan pada Gedung.
- SNI 03-6386-2000. Spesifikasi tingkat bunyi dan waktu dengung dalam bangunan gedung dan perumahan (kriteria desain yang direkomendasikan).
- SONG, G., HWANG, J., KIM, J., KIM, H., LEE, S., PARK, J., & CHO, G, 2024. Automatic Classification of Multi-Channel PSD Results Combining Unsupervised and Supervised Learning. *Nuclear Engineering and Technology*. <https://doi.org/10.1016/j.net.2024.11.033>
- SURAYA, S., SHOLEH, M., & LESTARI, U., 2023. Evaluation of Data Clustering Accuracy using K-Means Algorithm. *International Journal of Multidisciplinary Approach Research and Science*, 2(01), 385–396. <https://doi.org/10.59653/ijmars.v2i01.504>
- SURYADI, USEP TATANG., SARASWATI, SRI, 2020. Sistem Cerdas Pemantau Kenyamanan Ruang Kelas Berbasis Internet Of Things (Iot) Menggunakan Metode K-Means Pada Platform Thingspeak. *Jurnal Teknologi Informasi dan Komunikasi STMIK Subang*, Vol.15, No.1, April 2020. ISSN: 2252-4517.
- YAHYA, KURNIA BIN., MAHPUZ., 2019. Penggunaan Algoritma K-Means Untuk Menganalisa Pelanggan Potensial Pada Dealer SPS Motor Honda Lombok Timur Nusa Tenggara Barat. *Infotek: Jurnal Informatika dan Teknologi* Vol. 2, No. 2. pp. 109-118. <https://doi.org/10.29408/jit.v2i2.1447>

Halaman ini sengaja dikosongkan