

CLASSIFICATION OF ELDERLY HEALTH USING K-NEAREST NEIGHBOR COMPARISON, NAIVE BAYES AND DECISION TREE

KLASIFIKASI KESEHATAN LANSIA MENGGUNAKAN PERBANDINGAN METODE K-NEAREST NEIGHBOUR, NAIVE BAYES DAN DECISION TREE

Alviant Chandra Kusuma¹, Hari Soetanto²

^{1,2} Universitas Budi Luhur, Jl. Ciledug Raya, Kec. Pesanggrahan, Kota Jakarta Selatan, Indonesia
2111601072@student.budiluhur.ac.id¹, hari.soetanto@budiluhur.ac.id²

Abstract - Health and nutrition in the elderly play a crucial role in determining the quality of human resources, especially for the elderly themselves. The ageing process causes a decrease in the ability of body tissues to regenerate, making the elderly more vulnerable to infections and organ damage. Indonesia is currently experiencing an increase in the number of elderly, from 18 million people (7.56%) in 2010 to 25.9 million people (9.7%) in 2019, and is predicted to reach 48.2 million people (15.77%) in 2035. This study aims to determine the most effective algorithm for identifying the nutritional status of the elderly, by comparing three algorithms, namely Decision Tree, K-Nearest Neighbor (KNN), and Naïve Bayes. The methodology applied is CRISP-DM, and algorithm performance evaluation is carried out using the accuracy metric of the Confusion Matrix. The results showed that Decision Tree achieved the highest accuracy (95.55%), followed by Naïve Bayes (94.18%) and KNN (94.01%). The combination of algorithms provides optimal results because each algorithm can capture different patterns in the data, so the integration of the results can reduce errors and increase accuracy in the classification of the nutritional status of the elderly.

Keywords - Elderly Nutrition, K-NN, Naïve Bayes, Decision Tree.

Abstrak - Kesehatan dan gizi pada lansia memiliki peran krusial dalam menentukan kualitas sumber daya manusia, terutama bagi kelompok lansia itu sendiri. Proses penuaan menyebabkan penurunan kemampuan jaringan tubuh untuk regenerasi, sehingga lansia lebih rentan terhadap infeksi dan kerusakan organ. Indonesia saat ini mengalami peningkatan jumlah lansia, dari 18 juta jiwa (7,56%) pada tahun 2010 menjadi 25,9 juta jiwa (9,7%) pada tahun 2019, dan diprediksi akan mencapai 48,2 juta jiwa (15,77%) pada tahun 2035. Penelitian ini bertujuan untuk menentukan algoritma yang paling efektif dalam mengidentifikasi status gizi lansia, dengan membandingkan tiga algoritma, yaitu *Decision Tree*, *K-Nearest Neighbor* (KNN), dan *Naïve Bayes*. Metodologi yang diterapkan adalah CRISP-DM, dan evaluasi kinerja algoritma dilakukan menggunakan metrik akurasi dari *Confusion Matrix*. Hasil penelitian menunjukkan bahwa *Decision Tree* mencapai akurasi tertinggi (95,55%), diikuti oleh *Naïve Bayes* (94,18%) dan KNN (94,01%). Kombinasi algoritma memberikan hasil optimal karena setiap algoritma dapat menangkap pola yang berbeda dalam data, sehingga integrasi hasilnya dapat mengurangi kesalahan dan meningkatkan akurasi dalam klasifikasi status gizi lansia.

Kata Kunci - Gizi Lansia, K-NN, Naïve Bayes, Decision Tree.

I. PENDAHULUAN

Gizi masih menjadi persoalan nasional di Indonesia, di mana kelompok lansia termasuk yang paling sering terdampak. Lansia yang mengalami masalah gizi sejak usia muda berpotensi menghadapi gangguan pertumbuhan, penyakit, atau bahkan kematian. Salah satu upaya untuk menjaga kesehatan adalah dengan memastikan asupan makanan yang memadai untuk memenuhi kebutuhan tubuh. Selain itu, status gizi seseorang dapat digunakan sebagai tolok ukur untuk menilai kondisi kesehatannya[1]. Sejalan dengan permasalahan sebelumnya, Pulau Gili yang memiliki 1.285 lansia berusia 60 tahun ke atas, termasuk 157 orang yang berusia 100 tahun ke atas. Hasil penelitian menunjukkan bahwa hampir 90% lansia di pulau ini memiliki tingkat pendidikan rendah, sehingga pemahaman mereka tentang gizi seimbang masih terbatas[2]. Untuk mengatasi hal ini, salah satu solusi yang dapat dilakukan adalah mengadakan penyuluhan dan demonstrasi mengenai pemenuhan gizi seimbang bagi lansia. Kegiatan ini diharapkan dapat meningkatkan pengetahuan lansia serta mengubah pola hidup mereka, terutama dalam hal menjaga asupan gizi yang seimbang[3].

Orang lanjut usia mengalami penurunan fungsi tubuh, penyakit kronis, dan keterbatasan fisik. Setelah usia 40 tahun, metabolisme basal menurun sekitar 2% per tahun, dengan komposisi tubuh terdiri dari 30-40% zat padat dan 60-70% cairan. Beberapa masalah gizi yang umum dialami lansia antara lain penurunan nafsu makan, berat badan, perubahan indera pengecap, kesulitan mengunyah, menelan, konstipasi, serta keterbatasan dalam mengakses makanan. Mereka juga rentan terhadap kekurangan gizi, kelebihan berat badan, dan obesitas. Oleh karena itu, pemahaman tentang status gizi, kebutuhan nutrisi, dan edukasi menjadi faktor penting untuk memastikan pemenuhan gizi lansia[4]. Banyaknya lansia menghadapi masalah gizi akibat keterbatasan fisik dan penurunan metabolisme, memerlukan pendekatan berbasis teknologi untuk mendeteksi dan mengklasifikasikan status gizi mereka. Dengan demikian, sistem klasifikasi status gizi lansia yang akurat dan efektif sangat dibutuhkan untuk memastikan kondisi gizi mereka terdeteksi dengan tepat.

Penelitian terdahulu telah membahas berbagai metode dalam klasifikasi dan klusterisasi kesehatan, tetapi masih terdapat beberapa keterbatasan yang perlu diatasi. Studi pada referensi[5] hanya menggunakan *Naïve Bayes* dengan akurasi 91%, tanpa membandingkannya dengan metode lain. Selanjutnya studi pada referensi[6] menggunakan *K-Nearest Neighbor Clustering* untuk mengelompokkan status gizi balita dan lansia, namun tidak berfokus pada klasifikasi individu. Referensi[7] mengimplementasikan *Decision Tree C4.5* dalam klasifikasi hipertensi lansia dengan akurasi 90%, tetapi hanya terbatas pada penyakit tertentu. Referensi [8] membandingkan *Naïve Bayes* dan *Decision Tree* dalam klasifikasi diabetes, dengan *Decision Tree* unggul di 91.67%, tetapi penelitian ini tidak mempertimbangkan *K-Nearest Neighbour* (KNN). Sementara itu, referensi [9] menggunakan *Decision Tree* dan KNN dalam klasifikasi status kesehatan berdasarkan usia dan olahraga, namun penelitian ini tidak spesifik membahas kesehatan lansia dan tidak melibatkan *Naïve Bayes* sebagai perbandingan. Oleh karena itu, penelitian ini perlu dilakukan untuk menutup gap penelitian sebelumnya dengan membandingkan tiga metode sekaligus guna menemukan algoritma paling efektif dalam klasifikasi kesehatan lansia. Penelitian ini bertujuan untuk menemukan algoritma yang paling efektif dalam mengklasifikasikan status gizi lansia dengan menggunakan algoritma *Decision Tree*, *K-Nearest Neighbor*, dan *Naïve Bayes*.

II. SIGNIFIKASI STUDI

A. Penelitian Terdahulu

Dalam penelitian ini, ditemukan beberapa hasil studi sebelumnya yang menganalisis penggunaan algoritma *K-Nearest Neighbors* (KNN) dan *Naive Bayes* dan *Decision Tree*, yang dapat dirangkum sebagai berikut:

TABEL I
PENELITIAN TERDAHULU

Nomor	Penulis	Penelitian Terdahulu
1	Katina dkk.[5]	Klasifikasi Gizi Lansia Menggunakan Metode <i>Naive Bayes</i> . Algoritma <i>Naive Bayes</i> digunakan untuk mengumpulkan data dan memberikan rekomendasi terkait kondisi gizi lansia di Posbindu PTM. Informasi dan rekomendasi yang dihasilkan membantu Posbindu PTM dalam memberikan penanganan serta layanan langsung kepada pasien lansia, terutama bagi mereka yang mengalami malnutrisi. Berdasarkan pengujian terhadap 110 sampel data, algoritma ini mencapai tingkat akurasi sebesar 91%. Sistem Klasterisasi Algoritma <i>K-Nearest Neighbor</i> untuk Menentukan Perkembangan Gizi Balita dan Lansia Pada Posyandu Banjarsari.
2	Ramadan dkk.[6]	Hasil penelitian menunjukkan bahwa implementasi algoritma <i>K-Nearest Neighbor</i> memungkinkan pengelompokan data secara efektif, sehingga tenaga kesehatan dapat lebih mudah mengidentifikasi dan memantau status gizi individu. Sistem ini berhasil mencapai tingkat keberhasilan 85% dalam mengklasifikasikan status gizi lansia dan 90% untuk balita, menjadikannya solusi efektif dalam pemantauan status gizi di Posyandu.
3	Antikasari dkk. [7]	Implementasi <i>Decision Tree</i> C.4.5 pada Klasifikasi Penyakit Hipertensi di Posyandu Lansia Desa Balegondo Implementasi <i>Decision Tree</i> C.4.5 dalam klasifikasi penyakit hipertensi di Posyandu Desa Balegondo mencapai akurasi 90%. Sistem ini dapat menampilkan grade keparahan hipertensi, memungkinkan penanganan lebih intensif oleh petugas posyandu. Dengan demikian, kualitas layanan kesehatan bagi lansia meningkat.
4	Putra dkk. [8]	Perbandingan Klasifikasi antara <i>Naives Bayes</i> dan <i>Decision Tree</i> dalam Prediksi Penyakit Diabetes Tahap Awal Hasil penelitian menunjukkan bahwa algoritma <i>Decision Tree</i> unggul dalam mengidentifikasi pasien diabetes dengan akurasi 91,67%, dibandingkan <i>Naive Bayes</i> yang hanya mencapai 66,67%. Dengan demikian, <i>Decision Tree</i> lebih efektif dalam studi kasus ini.
5	Biyantoro & Prasetyo[9]	Penerapan <i>Decision Tree</i> untuk Klasifikasi Status Kesehatan dengan perbandingan KNN dan <i>Naive Bayes</i> . Hasil penelitian menunjukkan bahwa algoritma <i>Decision Tree</i> mencapai akurasi tertinggi 95,45% dalam memprediksi kesehatan berdasarkan usia dan intensitas olahraga. Intensitas olahraga lebih dari 2,5 jam per minggu menjadi faktor penting, sementara usia dan durasi olahraga mempengaruhi kategori kesehatan.

Berdasarkan tabel 1. penelitian terdahulu diketahui bahwa metode *Naive Bayes*, *Decision Tree*, dan KNN, dengan hasil yang bervariasi. *Naive Bayes* mencapai akurasi 91% dalam klasifikasi gizi lansia, KKN berhasil mengelompokkan data dengan tingkat keberhasilan 85% untuk lansia, dan *Decision Tree* menunjukkan performa tinggi dalam klasifikasi hipertensi (90%) serta diabetes (91,67%). Meskipun metode ini telah diterapkan, belum ada penelitian yang secara langsung

membandingkan *Naïve Bayes*, KNN, dan *Decision Tree* dalam klasifikasi kesehatan lansia, sehingga penelitian ini diperlukan untuk menentukan metode yang paling akurat.

B. Tinjauan Pustaka

1. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah proses utama dalam data mining, yaitu metode untuk mengekstraksi pengetahuan dari database guna mendukung proses pengambilan keputusan[10].

2. Data Selection

Seleksi data memerlukan data yang telah melalui proses penyaringan dan validasi. Dalam *Knowledge Discovery in Database* (KDD), tahap ini dilakukan sebelum data mining, dengan data terpilih disimpan terpisah dari basis data operasional. [10].

3. Pre-processing/Cleaning

Pembersihan data bertujuan untuk menghilangkan duplikasi dan data yang tidak relevan sebelum proses *data mining*. Proses ini mencakup pemeriksaan inkonsistensi, perbaikan kesalahan, dan pengayaan data sesuai kebutuhan penelitian. [11].

4. Transformation

Data yang telah difilter diubah menjadi matriks transformasi dengan pola teratur agar sesuai untuk proses data mining, melibatkan penyesuaian pola penyimpanan dalam basis data. Dalam definisi lain, transformasi adalah proses mengubah data terpilih ke dalam format yang siap digunakan untuk data mining, menghasilkan tabel integrasi sebagai *output* untuk tahap selanjutnya[11].

5. Data Mining

Data mining adalah proses mengidentifikasi pola dan membangun aturan untuk menghasilkan informasi yang mendukung pengambilan keputusan. Proses ini menggabungkan statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak pengetahuan yang tersembunyi dari database besar, dengan memilih metode yang sesuai dengan tujuan penelitian dan tahapan KDD [12].

6. Interpretation

Tahap interpretasi dalam *Knowledge Discovery in Database* (KDD) bertujuan untuk mengevaluasi pola informasi yang dihasilkan dari proses data mining agar mudah dipahami oleh pihak terkait. Pada tahap ini, dilakukan verifikasi untuk memastikan kesesuaian pola dengan hipotesis dan fakta yang telah ada [13].

7. Status Gizi Lansia

Status gizi yang optimal tercapai ketika asupan makanan dan minuman sesuai dengan kebutuhan tubuh. Pada lansia, masalah gizi seringkali berawal sejak usia muda dan diperparah oleh penurunan fungsi fisiologis, yang meningkatkan risiko penyakit degeneratif. Selain itu, faktor anatomis, fungsional, dan mental yang tidak stabil juga memengaruhi status gizi lansia[14].

8. K-Nearest Neighbors

K-Nearest Neighbor (*K-NN*) adalah algoritma pembelajaran mesin sederhana yang mengklasifikasikan objek berdasarkan kemiripan dengan tetangga terdekatnya. Metode ini merupakan pengembangan dari *Nearest Neighbor*, di mana objek baru diklasifikasikan berdasarkan mayoritas suara dari k tetangga terdekat[15].

9. *Nive Bayes*

Naïve Bayes adalah metode prediksi berbasis probabilitas yang mengasumsikan independensi antar atribut. Algoritma ini mengklasifikasikan data berdasarkan probabilitas suatu kelas dan mampu menghasilkan keputusan dari data pembelajaran[15].

10. *Crisp DM*

Cross-Industry Standard Process for Data Mining (CRISP-DM) adalah metodologi standar dalam penelitian *data mining* yang dikembangkan sejak 1996 melalui kolaborasi lima perusahaan: *Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation, dan OHRA*[16].

11. *Decious Tree*

Decision Tree adalah model berbentuk diagram pohon yang terdiri dari simpul internal (mewakili pengujian atribut), cabang (menunjukkan hasil pengujian), dan simpul daun (mewakili kelas atau distribusi kelas)[16].

12. *Confusion Matrix*

Confusion Matrix adalah alat yang digunakan untuk mengevaluasi kinerja model klasifikasi dalam *data mining* dengan membandingkan hasil prediksi algoritma terhadap klasifikasi yang sebenarnya[13].

13. *Black Box Testing*

Black Box Testing adalah metode pengujian yang memeriksa input dan output tanpa melihat struktur internal perangkat lunak, memastikan fungsionalitas sesuai persyaratan dan mendeteksi kesalahan[13].

14. *Machine learning*

Machine learning adalah cabang dari kecerdasan buatan yang mengembangkan algoritma untuk memungkinkan komputer belajar dari data. Tujuannya adalah secara otomatis mengenali pola kompleks dan membuat keputusan cerdas berdasarkan data tersebut[16]. Penerapan *Knowledge Discovery in Database (KDD)* dalam analisis gizi lansia dapat mengidentifikasi pola-pola yang memengaruhi status gizi kelompok usia ini. Melalui tahapan seleksi data, pembersihan, transformasi, dan data mining, peneliti dapat mengolah data relevan untuk mengungkap faktor-faktor seperti penurunan fungsi fisiologis dan masalah mental yang berkontribusi terhadap kekurangan gizi. Algoritma seperti *K-Nearest Neighbors* dan *Naïve Bayes* memungkinkan klasifikasi dan prediksi risiko gizi, sementara metodologi *Crisp-DM* memberikan kerangka kerja sistematis untuk penelitian. Evaluasi model menggunakan *Confusion Matrix* membantu menilai akurasi prediksi, mendukung perencanaan intervensi gizi yang lebih efektif. Dengan demikian, integrasi teknik data mining dapat meningkatkan pemahaman dan pengambilan keputusan terkait masalah gizi lansia.

C. *Metode Penelitian*

1. *Metode Penelitian*

Penelitian ini membandingkan metode *Decision Tree*, *K-Nearest Neighbor*, dan *Naïve Bayes* dalam klasifikasi status gizi lansia melalui tahapan pengumpulan data, preprocessing, implementasi algoritma, validasi model, dan analisis hasil. Data yang digunakan diproses melalui pembersihan (*data cleaning*), normalisasi, dan transformasi untuk memastikan kualitas data. Parameter utama untuk setiap algoritma disesuaikan, seperti pemilihan jumlah tetangga (k) pada KNN, kedalaman pohon keputusan pada *Decision Tree*, serta distribusi probabilitas pada *Naïve Bayes*. Model dievaluasi menggunakan *Confusion Matrix*, akurasi, *precision*, *recall*, dan *F1-score* guna menilai performa klasifikasi.

2. Sampel Penelitian

Dataset yang digunakan berasal dari *Kaggle*, terdiri dari 135 data lansia dengan variabel utama seperti umur, berat badan, tinggi badan, tinggi lutut, lingkaran lengan, lingkaran kepala, lingkaran dada, jaringan lunak, dan indeks massa tubuh (IMT). Kriteria inklusi meliputi individu berusia 60 tahun ke atas dengan data lengkap, sementara data yang tidak memenuhi kelengkapan variabel utama dikeluarkan dalam proses preprocessing.

3. Metode Pengumpulan Data

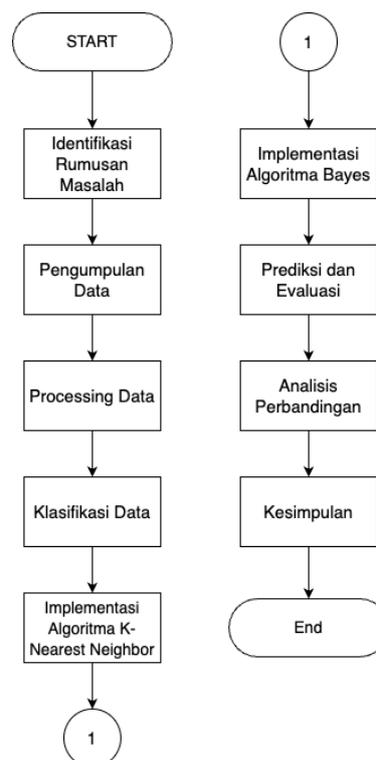
Pengumpulan data dilakukan melalui pengunduhan dataset dari *Kaggle*, kemudian data tersebut dianalisis untuk memastikan keterwakilan populasi lansia secara umum. Data yang tidak lengkap atau memiliki nilai ekstrem diatasi dengan teknik imputasi dan deteksi outlier. Distribusi data dianalisis menggunakan statistik deskriptif untuk memastikan bahwa setiap kelompok gizi lansia memiliki jumlah data yang cukup untuk pelatihan dan pengujian model.

4. Metode Analisis Data

Analisis data dilakukan menggunakan metode *Decision Tree*, KNN, dan *Naïve Bayes*, dengan pendekatan CRISP-DM sebagai kerangka kerja utama. Model dilatih dan diuji menggunakan teknik validasi silang (*cross-validation*) dengan rasio data latih dan uji sebesar 80:20. Pembagian data menjadi 80% untuk data latih (training set) dan 20% untuk data uji (test set) adalah praktik umum dalam pembelajaran mesin untuk memastikan model yang dibangun memiliki kemampuan generalisasi yang baik[17]. Performa model dievaluasi dengan *Confusion Matrix*, akurasi, *precision*, *recall*, dan *F1-score* guna memastikan efektivitas masing-masing metode dalam mengklasifikasikan status gizi lansia. Hasil dari setiap algoritma dibandingkan untuk menentukan metode yang paling optimal.

5. Langkah-Langkah Penelitian

Adapun beberapa langkah-langkah sistematis dalam Diagram Alur penelitian yaitu sebagai berikut:



Gambar 1. Diagram Alur Penelitian

Diagram alur penelitian ini menggambarkan tahapan sistematis dalam klasifikasi status gizi lansia menggunakan algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes*. Penelitian dimulai dengan identifikasi masalah dan pengumpulan data lansia dari sumber yang relevan. Data kemudian diproses melalui tahap *preprocessing*, termasuk pembersihan dan normalisasi, sebelum dilakukan klasifikasi menggunakan KNN dan *Naïve Bayes*. Hasil klasifikasi dievaluasi menggunakan metrik *Confusion Matrix* untuk menilai performa setiap algoritma. Selanjutnya, dilakukan analisis perbandingan guna menentukan metode yang lebih optimal dalam mengklasifikasikan status gizi lansia. Akhirnya, kesimpulan diambil berdasarkan hasil analisis sebagai dasar rekomendasi lebih lanjut.

Penelitian terdahulu menunjukkan variasi performa *Naïve Bayes* (91%), KNN (85%), dan *Decision Tree* (90-91,67%) dalam klasifikasi kesehatan, namun belum ada studi yang membandingkan ketiganya secara langsung untuk status gizi lansia. Penelitian ini mengisi gap tersebut dengan mengusulkan pendekatan sistematis menggunakan KDD dan CRISP-DM, serta membandingkan ketiga algoritma melalui evaluasi komprehensif (akurasi, *precision*, *recall*, *F1-score*) untuk menentukan metode paling efektif. Selanjutnya penelitian akan di lakukan sesuai langkah-langkah penelitian pada diagram alur penelitian. Penelitian ini diharapkan dapat mengisi gap penelitian terdahulu dan memberikan solusi teknologi inovatif bagi intervensi gizi lansia yang lebih tepat sasaran.

III. HASIL DAN PEMBAHASAN

A. Hasil

1. Pengumpulan Data

Penelitian ini mengevaluasi efektivitas metode *Decision Tree*, *K-Nearest Neighbor*, dan *Naïve Bayes* pada data gizi lansia. *Dataset* yang digunakan terdiri dari 135 data lansia yang diambil dari situs *Kaggle*, dengan kolom data meliputi umur, berat badan, tinggi badan, tinggi lutut, lingkar lengan, lingkar kepala, lingkar dada, jaringan lunak, dan indeks massa tubuh. Setelah melakukan *preprocessing* data, ketiga metode tersebut diterapkan untuk mengklasifikasikan status gizi lansia menjadi Gizi Kurang, Gizi Baik, dan Gizi Lebih. Evaluasi kinerja model dilakukan menggunakan metrik akurasi. Data awal untuk penelitian ini disajikan dalam tabel II berikut.

TABEL II
DATA AWAL

Umur	Berat Badan	Tinggi Badan	Tinggi Lutut	Lingkar Lengan Atas	Lingkar Kepala	Lingkar Dada	Jaringan Lunak	Imt	Status Gizi
65	62.5	165.2	48.5	26.5	56.2	94.5	15.8	22.9	Baik
72	48.2	158.5	46.2	21.8	55.1	85.2	11.5	19.2	Baik
68	85.4	172.3	51.2	32.4	57.5	102.8	24.2	28.7	Lebih
75	55.6	162.8	47.8	24.2	55.8	88.5	14.2	21.0	Baik
70	45.8	160.5	47.2	20.5	55.4	84.2	10.8	17.8	Kurang
65	62.5	165.2	48.5	26.5	56.2	94.5	15.8	22.9	Lebih

2. *Data Processing*

Data Lansia diatas tidak dapat langsung dipergunakan untuk dalam proses data mining, oleh karena itu diperlukan beberapa pemrosesan data terlebih dahulu sehingga data tersebut bisa di proses dengan baik dan benar:

a. *Data Cleaning*

Terdapat data yang tidak ada nilai atau NULL dan *blank* pada sebagian besar atribut dikarenakan tidak semua lansia diperiksa dan dicek dengan semua kategori atribut tersebut, proses dibawah dengan menggunakan *script python* untuk memastikan tidak ada *missing value*. Data cleaning disajikan pada tabel III *Missing Data* berikut.

TABEL III
MISSING DATA

Nama Atribut	Missing
umur	0
berat badan	0
tinggi badan	0
tinggi lutut	0
lingkar lengan	0
lingkar kepala	0
lingkar dada	0
jaringan lunak	0
indeks masa tubuh	0
Status	0

b. *Fitur Scalling*

Proses *fitur scalling* dengan metode *min-max* yang memungkinkan beberapa variabel memiliki nilai rentang yang sama dari angka 0 sampai angka 1.

TABEL III
FITUR SCALLING MIN MAX

Umur	Berat Badan	Tinggi Badan	Tinggi Lutut	Lingkar Lengan Atas	Lingkar Kepala	Lingkar Dada	Jaringan Lunak	Imt
0.167	0.422	0.495	0.486	0.514	0.500	0.557	0.396	0.457
0.458	0.104	0.157	0.157	0.192	0.176	0.179	0.126	0.138
0.292	0.931	0.854	0.871	0.918	0.882	0.894	0.925	0.957

c. *Pembagian data latih dan data uji*

Dalam penentuan data latih dan data uji ini menggunakan model seleksi *train test split* dimana pemilihan data dipilih secara acak dengan data uji sebesar 20% dan nilai random *state* 42.

TABEL IV
PROPORSI DATA

Penjelasan	Data Latih	Data Uji	Total
Proporsi	80%	20%	100%
Jumlah	108 data	27 data	135 data

3. *Pemodelan*

Setelah mendapatkan data latih dan data uji, selanjutnya dilakukan pemodelan menggunakan *Decision Tree*, *K-Nearest Neighbor* dan *Naive Bayes* ketiga model ini dibangun dan dikembangkan menggunakan bahasa pemrograman Python.

a. *Klasifikasi dengan dengan Naive Bayes, KNearest Neighbor dan Decision Tree*

Klasifikasi prediksi dengan *Naive Bayes*, *KNearest Neighbor* dan *Decision Tree* dilakukan dengan memisahkan data latih dan data uji seperti yang dilakukan diatas sesuai dengan atribut yang telah di tentukan adapun penelitian ini menggunakan perangkat lunak *streamlit* dan pemrograman *python*.

```

Initialization complete
Iteration 0, inertia 496.74911269684344.
Iteration 1, inertia 297.62147187254607.
Iteration 2, inertia 289.669190969634.
Iteration 3, inertia 287.8150982164851.
Converged at iteration 3: strict convergence.
Initialization complete
Iteration 0, inertia 517.2212751296597.
Iteration 1, inertia 326.4749399899402.
Iteration 2, inertia 316.34021822012926.
Iteration 3, inertia 306.63795218163716.
Iteration 4, inertia 304.777007019907.
Iteration 5, inertia 299.6511956618823.
Converged at iteration 5: strict convergence.
Initialization complete
Iteration 0, inertia 516.1176113017756.
Iteration 1, inertia 325.329511240583.
Iteration 2, inertia 308.729809626172.
Iteration 3, inertia 305.22251320220454.
Iteration 4, inertia 302.80864643030156.
Iteration 5, inertia 302.0404639921968.
Converged at iteration 5: strict convergence.
Initialization complete
Iteration 0, inertia 444.49117592213355.
Iteration 1, inertia 297.2783995483819.
    
```

Gambar 2. Proses Klasifikasi

d. *Pengujian Metode*

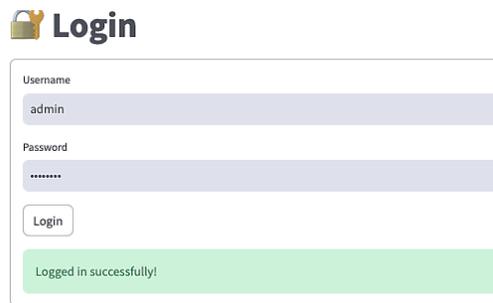
Dalam penelitian ini, klasifikasi menggunakan metode *Naive Bayes*, *K-Nearest Neighbor* (KNN), dan *Decision Tree* diterapkan menggunakan *Python* melalui tahapan persiapan data, *upload* data lansia, pembagian data (80% untuk pelatihan dan 20% untuk pengujian), serta validasi silang (*cross-validation*). Hasil analisis menunjukkan performa masing-masing model sebagai berikut.

TABEL V
AKURASI PERBANDINGAN MODEL

Model	Akurasi (%)	Precision (Baik) (%)	Precision (Lebih) (%)	Precision (Kurang) (%)	Recall (Baik) (%)	Recall (Lebih) (%)	Recall (Kurang) (%)	F1-Score (Baik) (%)	F1-Score (Lebih) (%)	F1-Score (Kurang) (%)
Naive Bayes	94,18	98,33	93,94	66,67	96,63	93,94	50	96,63	93,94	57,14
KNN	94,01	95,31	93,94	80	95	92	70	95,15	92,96	74,07
Decision Tree	95,55	94,12	98,39	80	95	96	75	94,56	96,94	77,78

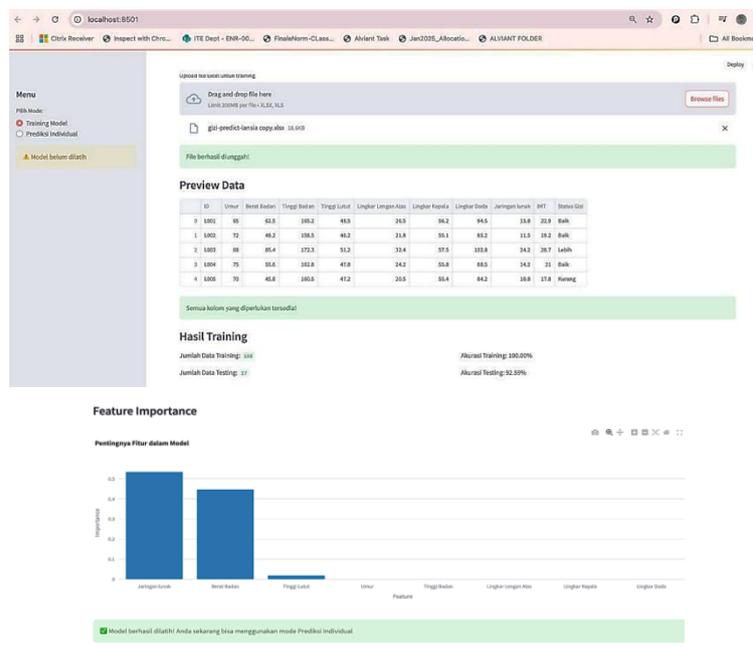
e. Aplikasi Prototype

Sebelum menggunakan sistem, pengguna harus login dengan memasukkan username dan password. Jika berhasil, akses aplikasi diberikan, namun jika gagal, muncul popup "Login Failed". Berikut tampilan login aplikasi pada gambar 3.



Gambar 3. Tampilan Login Aplikasi

Setelah berhasil login, aplikasi akan menampilkan prototype klasifikasi kesehatan lansia menggunakan Decision Tree. Halaman ini mencakup fitur hasil training, seperti jumlah data training, akurasi training, jumlah data testing, serta validasi untuk mengevaluasi keakuratan model dalam mengklasifikasikan data.



Gambar 4. Tampilan Hasil Training

Berikut adalah contoh hasil prediksi status gizi lansia, yang terdiri dari gizi kurang, gizi baik, dan gizi lebih.



Gambar 5. Tampilan Hasil prediksi gizi dengan aplikasi (a) gizi kurang, (b) gizi baik dan (a) gizi lebih baik.

B. Pembahasan

Hasil penelitian ini menunjukkan bahwa tujuan untuk membandingkan tiga model klasifikasi *Decision Tree*, *K-Nearest Neighbors* (KNN), dan *Naive Bayes* dalam menentukan status gizi lansia telah tercapai. Dengan akurasi tertinggi sebesar 95.55%, *Decision Tree* mengungguli KNN (94.01%) dan *Naive Bayes* (94.18%), yang menunjukkan kemampuannya dalam menangani pola *non-linear* dan hubungan antar fitur dalam data. Penelitian ini berhasil mengisi gap yang ada dalam literatur sebelumnya, di mana studi-studi terdahulu seperti yang tercantum dalam referensi [5] hanya menggunakan *Naive Bayes* dengan akurasi 91% tanpa perbandingan dengan metode lain, dan referensi[6] yang menggunakan KNN tidak fokus pada klasifikasi individu. Selain itu, penelitian pada referensi[7] dan[8] terbatas pada penyakit tertentu dan tidak mempertimbangkan KNN, sedangkan referensi[9] tidak spesifik membahas kesehatan lansia. Dengan membandingkan ketiga metode secara bersamaan, penelitian ini memberikan wawasan baru mengenai efektivitas masing-masing algoritma dalam klasifikasi status gizi lansia.

Namun, perlu dicatat bahwa potensi *overfitting* atau *underfitting* pada model yang dibangun harus diperhatikan. Meskipun *Decision Tree* menunjukkan akurasi tinggi, risiko *overfitting* dapat muncul jika model terlalu rumit. Sebaliknya, model yang terlalu sederhana dapat mengalami *underfitting*, mengakibatkan kinerja buruk pada data baru[17]. Oleh karena itu, validasi silang yang lebih mendalam dan pengujian pada dataset yang berbeda diperlukan untuk memastikan bahwa model dapat digeneralisasi dengan baik di luar *dataset* yang digunakan.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa *Decision Tree* memiliki akurasi tertinggi (95,55%) dalam mengklasifikasikan status gizi lansia, mengungguli KNN dan *Naive Bayes*, karena kemampuannya menangani pola *non-linear* dan hubungan kompleks antar fitur. KNN bergantung pada pemilihan parameter k , sementara *Naive Bayes* terbatas oleh asumsi independensi fitur yang sering tidak terpenuhi. Keterbatasan penelitian meliputi ukuran dataset yang kecil dan potensi bias dalam pemilihan data. Kombinasi algoritma memberikan hasil optimal karena setiap algoritma dapat menangkap pola yang berbeda dalam data, sehingga integrasi hasilnya dapat mengurangi kesalahan dan meningkatkan akurasi dalam klasifikasi status gizi lansia. Untuk penelitian selanjutnya, disarankan menggunakan teknik *ensemble learning* seperti *Random Forest*, mengeksplorasi metode lain seperti SVM, serta mengoptimalkan *Decision Tree* untuk menghindari *overfitting*. Hasil penelitian ini dapat diintegrasikan ke dalam sistem kesehatan lansia untuk mendukung penilaian dan intervensi gizi yang lebih efektif.

REFERENSI

- [1] Rumlah S. Masalah Sosial dan Solusi dalam Menghadapi Fenomena Stunting pada Anak. *J Pendidik Sej Sej FKIP Univ Jambi*. 2022;1(3):83–91.
- [2] Gamar I, Pratiwi D, Yasin Z, Suprayitno E. Pemenuhan Gizi Seimbang Bagi Lansia Di Pulau Gili Iyang Kecamatan Dungkek. *J Abdiraja*. 2021;2(1):12–8.
- [3] Sunarti A, Graha K, Palu A. Konsep Pelayanan Kesehatan pada Kelompok Rentan, Pus, Lansia, Bayi dan Balita pada Masa Covid-19 Concepts of Health Services in Vulnerable Groups, Couples of Childbearing Age, Elderly, Infants and Toddlers in the Time of Covid-19. *MPPK Media Publ Penelit Kebidanan*. 2020;3(2):41–50.
- [4] Yosephin B. Tuntunan Praktis Menghitung Kebutuhan Gizi. Yogyakarta: CV. Andi Offset; 2020.
- [5] Katina, Hatina AA, Rismayati R, Rahmiati BF, Fatimatu Zahra, Amrullah R, et al. Klasifikasi Gizi Lansia menggunakan Metode Naïve Bayes Classifier. *JTIM J Teknol Inf dan Multimed*. 2024;6(2):84–96.
- [6] Ramadhan DG, Kasih P, Irawan RH. Sistem Klasterisasi Algoritma K-means Untuk Menentukan Perkembangan Gizi Balita Dan Lansia Pada Posyandu Banjarsari. In: Seminar Nasional Inovasi Teknologi UN PGRI Kediri, Kediri: UN PGRI Kediri; 2022. p. 111–6.
- [7] Antikasari EW, Mustikasari D, Litanianda Y. Implementasi Decision Tree C.4.5 pada Klasifikasi Penyakit Hipertensi di Posyandu Lansia Desa Balegondo. *KOMPUTEK J Tek Univ Muhammadiyah Ponorogo ISSN [Internet]*. 2021;1(1):1–14.
- [8] Putra AW, Kosumo K, Sitho A, Ratu R, Mujayanto RR. Perbandingan Klasifikasi antara Naives Bayes dan Decision Tree dalam Prediksi Penyakit Diabetes Tahap Awal. *J Ilmu Komput*. 2021;17(1):46–54. .
- [9] Biyantoro AS, Prasetyo B. Penerapan Decision Tree untuk Klasifikasi Status Kesehatan dengan perbandingan KNN dan Naive Bayes. *IJIRSE Indones J Inform Res Softw Eng*. 2024;4(1):47–55.
- [10] Harahap B. Implementasi Data Mining dalam Pemilihan Laptop Berbasis Algoritma C4.5 Pada Software WEKA. *J Minfo Polgan*. 2023;12(2):1065–73.
- [11] Supoyo A, Prasetyaningrum PT. Analisis Data Mining Untuk Memprediksi Lama Perawatan Pasien Covid-19 di DIY. *Bianglala Inform*. 2022;10(1):21–9.
- [12] Anastassia S, Khari A, Haqqi A, Zili A. Learning Analytics dan Educational Data Mining pada Data Pendidikan Learning Analytics dan Educational Data Mining pada Data Pendidikan. *JRPMS (Jurnal Ris Pembelajaran Mat Sekolah)*. 2023;6(March 2022):12–20.
- [13] Adzy LB, Pambudi A, Sukabumi UM, Bantuan P, Jaminan I, Sukabumi SK. Algoritma Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Iuran Jaminan Kesehatan Pemerintah Daerah Kabupaten Sukabumi. *J Mnemon*. 2023;6(1):1–10.
- [14] Akbar F, Hamsah IA, M AM. Gambaran Nutrisi Lansia Di Desa Banua Baru. *J Ilm Kesehat Sandi Husada*. 2020;9(1):1–7.
- [15] Saputro MB. Comparison of Naive Bayes Classifier and K-Nearest Neighbor Algorithms with Information Gain and Adaptive Boosting for Sentiment Analysis of Spotify App Reviews. *Recursive J Informatics*. 2024;2(1):37–44.
- [16] Agung A, Daniswara A, Nuryana IKD. Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *JINACS J Informatics Comput Sci*. 2023;05(01):97–100.
- [17] Divva G, Zulma M, Chamidah N. Perbandingan Metode Klasifikasi Naive Bayes, Decision Tree Dan K- Nearest Neighbor Pada Data Log Firewall. In: Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA). Jakarta; 2021. p. 679–88.