



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 3, No. 1, 2026

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 888–898

Penerapan Algoritma C4.5 untuk Klasifikasi Gender Berdasarkan Ciri Fisik pada Dataset Gender Classification v7

Vina Meylani Sari, Hasbi Firmansyah, Wahyu Asriyani

Universitas Pancasakti Tegal

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

DOI : <https://doi.org/10.32672/mister.v3i1.4014>

How to Cite this Article

APA : Sari, V. M., Hasbi Firmansyah, & Wahyu Asriyani. (2025). Penerapan Algoritma C4.5 untuk Klasifikasi Gender Berdasarkan Ciri Fisik pada Dataset Gender Classification v7. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 3(1), 888 – 898.
<https://doi.org/10.32672/mister.v3i1.4014>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



Penerapan Algoritma C4.5 untuk Klasifikasi Gender Berdasarkan Ciri Fisik pada *Dataset Gender Classification v7*

Vina Meylani Sari¹, Hasbi Firmansyah², Wahyu Asriyani³

Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal,
Kota Tegal, Indonesia^{1,2}

Program Studi Pendidikan Bahasa dan Sastra Indonesia, Fakultas Keguruan dan Ilmu Pendidikan,
Universitas Pancasakti Tegal, Tegal, Indonesia³

*Email : vinameylanis@gmail.com, hasbifirmansyah@upstegal.ac.id, asriyani1409@gmail.com

Diterima: 09-12-2025

| Disetujui: 19-12-2025

| Diterbitkan: 21-12-2025

ABSTRACT

Gender classification based on physical attributes is one of the applications of data mining techniques widely used to support identification systems and demographic analysis. This study aims to apply the Decision Tree C4.5 algorithm for gender classification based on physical characteristics using the Gender Classification v7 Dataset. The C4.5 algorithm was selected due to its ability to construct an interpretable decision tree model and effectively handle both numerical and categorical attributes. The research stages include dataset retrieval, data preprocessing, label attribute assignment, data splitting into training and testing sets, model training using the C4.5 algorithm, and performance evaluation using accuracy, precision, and recall metrics. The experimental process was conducted using Altair AI Studio 2026.0.1, with a training data proportion of 70% and testing data proportion of 30%. The experimental results show that the C4.5 algorithm achieved an accuracy of 97.20%. Based on the confusion matrix, the Male class obtained a recall value of 95.97% and a precision value of 98.35%, while the Female class achieved a recall value of 98.41% and a precision value of 96.12%. These results indicate that physical attributes contribute significantly to gender differentiation and demonstrate that the C4.5 algorithm is an effective and reliable method for physical attribute-based gender classification.

Keywords: C4.5, decision tree, gender classification, data mining, gender classification v7.

ABSTRAK

Klasifikasi gender berdasarkan data fisik merupakan salah satu penerapan teknik data mining yang banyak digunakan untuk mendukung sistem identifikasi dan analisis demografis. Penelitian ini bertujuan untuk menerapkan algoritma Decision Tree C4.5 dalam melakukan klasifikasi gender berdasarkan ciri fisik pada Dataset Gender Classification v7. Algoritma C4.5 dipilih karena kemampuannya dalam membangun model pohon keputusan yang mudah dipahami serta efektif dalam menangani atribut numerik dan kategorikal. Tahapan penelitian meliputi pengambilan dataset, preprocessing data, penentuan atribut label, pembagian data menjadi data latih dan data uji, pelatihan model menggunakan algoritma C4.5, serta evaluasi performa model menggunakan metrik akurasi, precision, dan recall. Proses eksperimen dilakukan menggunakan Altair AI Studio 2026.0.1 dengan proporsi data latih sebesar 70% dan data uji sebesar 30%. Hasil pengujian menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan gender dengan tingkat akurasi sebesar 97,20%. Berdasarkan confusion matrix, kelas Male memperoleh nilai recall sebesar 95,97% dan precision sebesar 98,35%, sedangkan kelas Female memperoleh nilai recall sebesar 98,41% dan precision sebesar 96,12%. Hasil ini menunjukkan bahwa atribut fisik berkontribusi signifikan terhadap diferensiasi gender dan membuktikan bahwa algoritma C4.5 merupakan metode yang efektif dan reliabel untuk klasifikasi gender berbasis atribut fisik.

96,12%. Tingginya nilai akurasi, precision, dan recall tersebut menunjukkan bahwa ciri fisik memiliki kontribusi yang signifikan dalam membedakan gender, serta membuktikan bahwa algoritma C4.5 merupakan metode yang efektif dan layak digunakan untuk klasifikasi gender berbasis data fisik.

Katakunci: C4.5, decision tree, klasifikasi gender, data mining, gender classification v7.

PENDAHULUAN

Perkembangan teknologi informasi telah mendorong pemanfaatan data dalam berbagai bidang, termasuk dalam proses pengambilan keputusan berbasis kecerdasan buatan. Salah satu teknik yang banyak digunakan adalah data mining, yaitu proses menggali pola tersembunyi dari kumpulan data dengan menggunakan metode statistik, matematika, dan algoritma pembelajaran mesin. Salah satu cabang utama dalam data mining adalah klasifikasi, yang bertujuan untuk memetakan data ke dalam label tertentu berdasarkan pola atau karakteristik yang ditemukan dalam data tersebut. Klasifikasi menjadi penting ketika suatu sistem membutuhkan kemampuan untuk membedakan objek atau individu ke dalam kategori tertentu berdasarkan fitur yang tersedia.

Dalam era perkembangan teknologi informasi, data menjadi elemen penting dalam mendukung proses analisis dan pengambilan keputusan. Salah satu teknik yang memiliki peran signifikan dalam pengolahan data adalah klasifikasi, yaitu proses mengelompokkan data berdasarkan karakteristik tertentu. Metode klasifikasi telah banyak digunakan dalam berbagai bidang seperti kesehatan, perdagangan, dan demografi karena kemampuannya dalam mengidentifikasi pola dari data yang kompleks (Science, 2022).

Gender merupakan salah satu variabel demografis yang sering digunakan dalam analisis data karena perannya yang penting dalam penelitian sosial, kesehatan, maupun pemasaran. Informasi mengenai gender sering kali dapat diprediksi melalui karakteristik fisik seperti tinggi dan berat badan. Meskipun teknologi modern telah menghasilkan algoritma canggih seperti deep learning, metode seperti pohon keputusan tetap relevan karena menawarkan interpretasi yang mudah dipahami serta proses perhitungan yang efisien (Siregar, 2024).

Algoritma C4.5 merupakan salah satu algoritma pohon keputusan yang paling banyak diterapkan karena kemampuannya dalam menangani atribut numerik dan kategorik. Algoritma ini menggunakan konsep gain ratio untuk memilih atribut terbaik dalam membangun struktur pohon keputusan, sehingga menghasilkan model yang lebih stabil dan akurat. Kemampuannya dalam membentuk aturan keputusan yang mudah dianalisis menjadikannya alat yang efektif dalam penelitian klasifikasi, termasuk klasifikasi berbasis ciri fisik (Kunci, 2023)

Penelitian antropometri menunjukkan adanya perbedaan yang konsisten antara laki-laki dan perempuan dalam hal komposisi tubuh, ukuran fisik, dan massa otot. Perbedaan tersebut tidak hanya terlihat pada usia remaja dan dewasa, tetapi juga pada kelompok usia tertentu seperti atlet muda yang menunjukkan perbedaan signifikan dalam struktur tubuh berdasarkan gender. Temuan seperti ini mendukung penggunaan atribut fisik sebagai indikator potensial dalam klasifikasi gender (Podstawski et al., 2022).

Selain itu, studi pada populasi lansia menunjukkan bahwa perbedaan tinggi dan berat badan antara laki-laki dan perempuan tetap konsisten meskipun terjadi perubahan fisiologis akibat penuaan. Hal ini menegaskan bahwa ciri fisik dapat menjadi dasar yang cukup kuat untuk membedakan gender dalam berbagai kondisi dan kelompok usia. Dengan demikian, penggunaan data fisik sederhana seperti tinggi dan berat badan dapat dioptimalkan dalam model klasifikasi berbasis algoritma (Su et al., 2023).

penelitian ini bertujuan menerapkan algoritma C4.5 untuk mengklasifikasikan gender menggunakan Dataset Gender Classification v7 yang berisi data tinggi dan berat badan. Penelitian ini mengukur performa model melalui metrik evaluasi seperti akurasi, presisi, dan confusion matrix guna mengetahui efektivitas algoritma dalam memetakan pola dari data numerik sederhana. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan metode klasifikasi berbasis ciri fisik serta menjadi referensi bagi penelitian sejenis di masa mendatang (Dan et al., 2022).

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk menganalisis kemampuan algoritma Decision Tree dalam melakukan klasifikasi Gender. Proses penelitian dilakukan melalui beberapa tahapan, mulai dari pengumpulan dataset, preprocessing, pembagian data, pembangunan model, hingga evaluasi performa (Amaliyah & Fatah, 2024). Seluruh proses pengolahan data dilakukan menggunakan perangkat lunak RapidMiner Studio karena aplikasi ini memiliki antarmuka yang memudahkan penyusunan alur analisis tanpa memerlukan pemrograman langsung (Studi et al., n.d.).



Gambar 1. Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan, yang diilustrasikan pada diagram alur berikut:

1. Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini merupakan dataset numerik dan kategorikal yang berisi kumpulan data ciri fisik manusia yang digunakan untuk keperluan klasifikasi gender. Dataset tersebut adalah Gender Classification v7, yang telah tersedia dalam bentuk data terstruktur sehingga dapat langsung digunakan pada proses data mining dan machine learning tanpa memerlukan ekstraksi fitur tambahan dari citra atau sinyal mentah. Dataset Gender Classification v7 terdiri dari atribut-atribut fisik yang merepresentasikan karakteristik antropometri dan morfologi wajah, antara lain panjang rambut, ukuran dahi, bentuk hidung, ketebalan bibir, serta jarak antara hidung dan bibir. Atribut-atribut tersebut mencakup fitur numerik seperti forehead width dan forehead height, serta fitur kategorikal biner seperti long hair, nose wide, nose long, lips thin, dan distance nose to lip long. Seluruh fitur ini digunakan sebagai variabel independen dalam proses klasifikasi. Setiap baris data pada dataset merepresentasikan satu individu, sedangkan atribut gender digunakan sebagai label atau kelas target dengan dua kategori, yaitu Male dan Female. Dataset ini memiliki jumlah data yang cukup besar dan distribusi kelas yang relatif seimbang, sehingga mampu mendukung proses pelatihan dan pengujian model klasifikasi secara optimal. Karena dataset telah berbentuk nilai fitur dan tidak mengandung missing value, proses pengumpulan data difokuskan pada tahap pemilihan dataset yang relevan, pengimporan data ke dalam Altair AI Studio, serta

pemeriksaan struktur dan tipe data. Dengan karakteristik tersebut, dataset bersifat terstruktur, konsisten, dan mudah diproses dalam sistem klasifikasi berbasis algoritma Decision Tree C4.5.

2. Preprocessing

Tahap preprocessing data dilakukan untuk memastikan bahwa dataset Gender Classification v7 berada dalam kondisi yang layak sebelum digunakan dalam proses klasifikasi. Pada tahap ini dilakukan pemeriksaan terhadap keberadaan nilai kosong (missing value), data duplikat, serta kesesuaian tipe data pada setiap atribut. Hasil pemeriksaan menunjukkan bahwa dataset tidak mengandung nilai kosong dan tidak terdapat data duplikat, sehingga tidak diperlukan proses pembersihan data lebih lanjut. Atribut numerik seperti forehead width dan forehead height dipastikan bertipe numerik, sedangkan atribut kategorikal biner seperti long hair, nose wide, dan lips thin dipastikan memiliki nilai diskrit yang konsisten agar dapat diproses dengan baik oleh algoritma C4.5. Normalisasi atau standarisasi data tidak dilakukan karena algoritma Decision Tree C4.5 tidak sensitif terhadap perbedaan skala data dan mampu menangani atribut numerik secara langsung. Dengan demikian, proses preprocessing difokuskan pada penyiapan struktur dan kualitas data sehingga dataset bersifat terstruktur, konsisten, dan siap digunakan pada tahap pemodelan.

3. Pembagian Data

Dataset dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Proses pembagian data dilakukan menggunakan operator split data dengan proporsi 90% data untuk pelatihan dan 10% untuk pengujian. Pembagian ini dipilih agar model memperoleh data pelatihan yang cukup tanpa mengurangi representasi data pada tahap evaluasi. Selain split data, penelitian ini juga menerapkan accuracy dan classification error untuk memastikan hasil evaluasi lebih stabil dan tidak bias terhadap satu pembagian data tertentu (Algoritma et al., 2020)

4. Pembangunan Model

Pada tahap pembangunan model, penelitian ini menggunakan algoritma Decision Tree C4.5 yang membangun pohon keputusan berdasarkan pemilihan atribut terbaik menggunakan information gain dan gain ratio untuk mengklasifikasikan gender menjadi kelas Male dan Female. Model dibangun menggunakan operator Decision Tree pada Altair AI Studio Educational 2026 dengan konfigurasi parameter standar, termasuk pengaturan kedalaman pohon dan pruning untuk mencegah overfitting. Pelatihan dilakukan menggunakan 70% data latih dan 30% data uji, kemudian model divisualisasikan dalam bentuk pohon keputusan untuk analisis hasil klasifikasi

5. Evaluasi Performa

kinerja model klasifikasi diukur menggunakan beberapa metrik evaluasi yang umum digunakan dalam penelitian machine learning, khususnya untuk permasalahan klasifikasi biner. Evaluasi dilakukan melalui operator Performance (Classification) pada RapidMiner Studio. Metrik utama yang digunakan adalah akurasi, yaitu persentase prediksi model yang sesuai dengan kelas sebenarnya.

HASIL DAN PEMBAHASAN

Pengumpulan Dataset

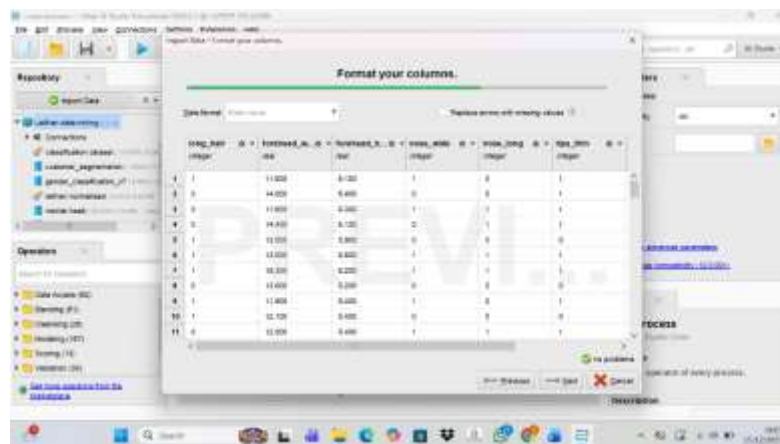
Penelitian ini bertujuan untuk mengklasifikasikan gender berdasarkan ciri fisik menggunakan algoritma C4.5 pada dataset Gender Classification v7. Proses eksperimen dilakukan menggunakan Altair

AI Studio 2026 dengan tahapan pengambilan data, praproses, pembagian data, pelatihan model, serta evaluasi performa.

Table 1. Dataset Gender Classification v7

No	long_hair	forehead_width	forehead_height	nose_wide	nose_long	...	gender
1	1	11.8	6.1	1	0		Male
2	0	14	5.4	0	0		Female
3	0	11.8	6.3	1	1		Male
4	0	14.4	6.1	0	1		Male
5	1	13.5	5.9	0	0		Female
6	1	13	6.8	1	1		Male
7	1	15.3	6.2	1	1		Male
8	0	13	5.2	0	0		Female
...	...	...	...	...	...	...	...
5002	1	15.4	5.4	1	1	...	Male

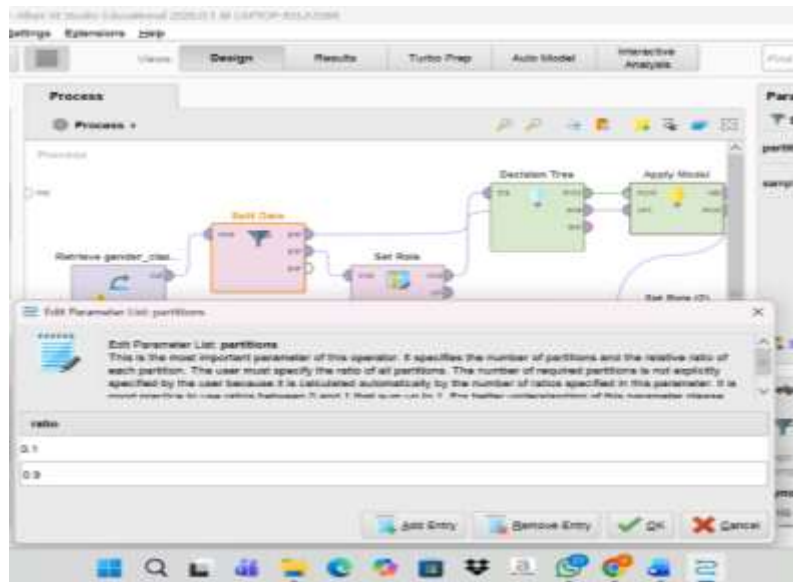
Preprocessing



Gambar 2. Preprocessing Dataset

Setelah dataset berhasil dikumpulkan, tahap berikutnya adalah melakukan preprocessing untuk memastikan bahwa data berada dalam kondisi optimal sebelum digunakan dalam proses pemodelan. Pada tahap ini, seluruh atribut numerik diperiksa secara menyeluruh menggunakan RapidMiner untuk memastikan tidak terdapat nilai hilang (*missing values*), duplikasi data, ataupun ketidakkonsistenan format yang dapat mengganggu proses analisis. Pemeriksaan ini dilakukan untuk mengidentifikasi potensi anomali yang dapat memengaruhi kinerja algoritma pada tahap pelatihan. Hasil pengecekan menunjukkan bahwa dataset Brain Tumor yang digunakan tidak mengandung nilai kosong sehingga tidak diperlukan proses imputasi atau pengisian nilai. Selain itu, atribut yang tersedia berada dalam format numerik yang seragam sehingga cocok digunakan sebagai input algoritma klasifikasi tanpa memerlukan proses konversi tipe data. Tahap preprocessing ini menjadi krusial karena memastikan bahwa dataset telah memenuhi standar kualitas data sebelum melangkah ke tahap normalisasi, pembagian dataset, dan pembangunan model klasifikasi yang lebih kompleks.

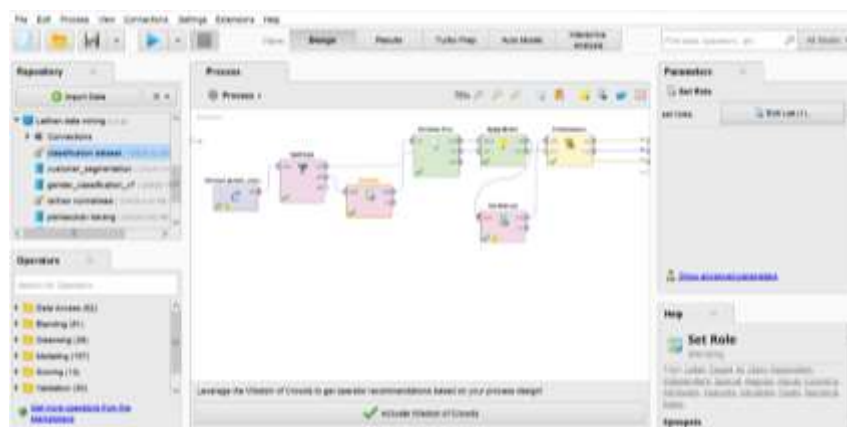
Pembagian Dataset



Gambar 3. Pembagian Data

Proses pembagian dataset menjadi data pelatihan dan data pengujian merupakan tahap yang sangat krusial dalam penelitian ini, karena menentukan sejauh mana model mampu mempelajari pola yang terdapat pada data sekaligus mengukur kemampuan generalisasinya terhadap data baru. Pembagian dilakukan menggunakan operator Split Data pada RapidMiner dengan rasio 90% data pelatihan dan 10% data pengujian. Proporsi ini dipilih karena dataset memiliki jumlah instance yang cukup besar, sehingga alokasi mayoritas data pada tahap pelatihan memungkinkan algoritma Decision Tree mempelajari karakteristik distribusi fitur secara lebih mendalam. Sementara itu, proporsi 10% data pengujian tetap dipandang memadai untuk memberikan gambaran performa model secara objektif, karena masih mampu merepresentasikan karakteristik kelas pada dataset secara proporsional.

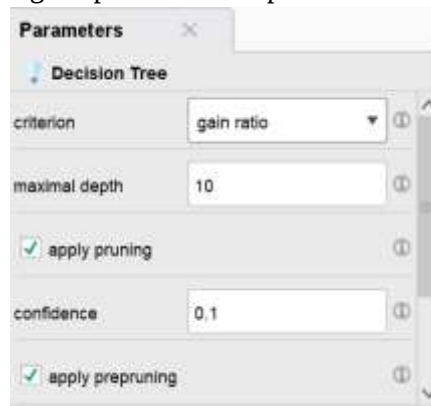
Pembangunan Model



Gambar 4. Pembangunan Model

Workflow pembangunan model yang terdiri dari beberapa operator inti seperti Retrieve, Select

Attributes, Split data, Set Role, Decision Tree, Apply Model, dan Performance. Workflow ini menggambarkan rangkaian proses otomatis yang saling terhubung mulai dari memanggil dataset hingga menghasilkan nilai akurasi dan confusion matrix. Struktur workflow yang jelas memudahkan replikasi penelitian, memastikan bahwa setiap langkah pemrosesan dapat ditelusuri dan divalidasi oleh peneliti lain.



Gambar 5. Parameter 1

Konfigurasi parameter pada algoritma Decision Tree. Parameter seperti *criterion* (Information Gain), *maximal depth*, *confidence*, dan *pruning* diatur berdasarkan default RapidMiner dengan aktivasi pruning untuk mengurangi risiko overfitting. Visualisasi parameter ini menjelaskan bagaimana model melakukan pemilihan atribut pemisah terbaik dan bagaimana mekanisme pemangkasan (pruning) bekerja untuk menghasilkan pohon keputusan yang optimal. Pengaturan parameter yang tepat merupakan faktor penting dalam meningkatkan performa dan interpretabilitas model.

Evaluasi Model

Evaluasi performa model dilakukan menggunakan operator *Performance (Classification)*. Metrik evaluasi yang digunakan mencakup akurasi, precision, recall, F1-score. Seluruh metrik ini digunakan untuk memberikan gambaran menyeluruh terhadap kemampuan model dalam mengklasifikasikan kedua kelas.

accuracy: 97.20%			
	true Male	true Female	class precision
pred. Male	238	4	98.38%
pred. Female	10	248	98.12%
class recall	95.97%	98.41%	

Gambar 6. Hasil Akurasi

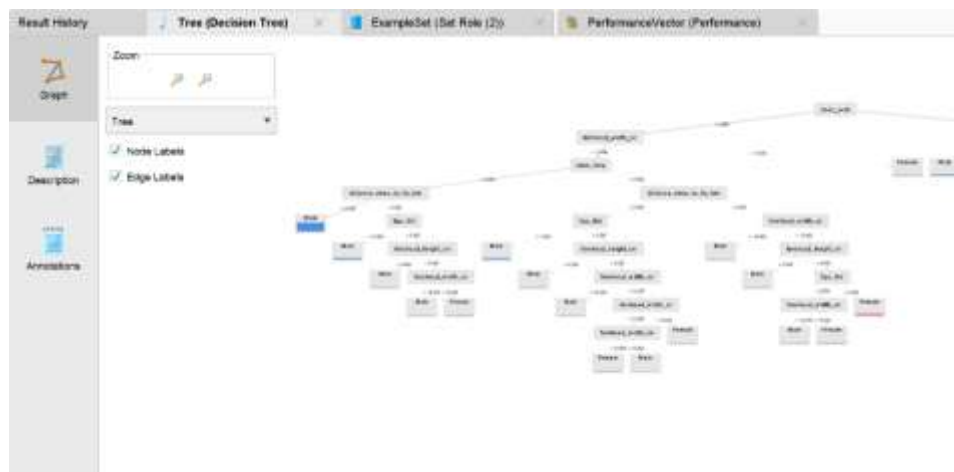
Hasil evaluasi menunjukkan bahwa model klasifikasi gender menggunakan algoritma Decision Tree C4.5 memperoleh nilai akurasi sebesar 97,20%. Berdasarkan nilai akurasi tersebut, dapat disimpulkan bahwa model mampu mengklasifikasikan sebagian besar data uji dengan benar. Model berhasil mengklasifikasikan 238 data berlabel Male secara tepat, dengan 4 data Female yang salah diklasifikasikan sebagai Male, serta 248 data Female yang berhasil diklasifikasikan dengan benar, dan 10 data Male yang salah diklasifikasikan sebagai Female.

Nilai precision dan recall yang diperoleh untuk kedua kelas juga tergolong tinggi. Pada kelas Male, precision mencapai 98,35% dengan recall sebesar 95,97%, sedangkan pada kelas Female, precision mencapai 96,12% dan recall sebesar 98,41%. Hasil ini menunjukkan bahwa model memiliki tingkat ketepatan dan sensitivitas yang baik dalam membedakan kedua kelas gender, baik dalam meminimalkan kesalahan prediksi positif maupun dalam mengenali data yang sebenarnya termasuk ke dalam kelas tersebut.



Gambar 7. Hasil Performance Vector

Hasil evaluasi performa model yang ditampilkan melalui Performance Vector menunjukkan bahwa algoritma Decision Tree C4.5 mampu menghasilkan kinerja klasifikasi yang sangat baik dalam membedakan gender berdasarkan ciri fisik pada dataset Gender Classification v7. Model mencapai tingkat akurasi sebesar 97,20%, yang mengindikasikan bahwa sebagian besar hasil prediksi telah sesuai dengan label sebenarnya pada data uji, dengan tingkat kesalahan klasifikasi yang relatif rendah. Berdasarkan confusion matrix, kesalahan klasifikasi pada kedua kelas tergolong kecil dan seimbang, yang menunjukkan bahwa model tidak bias terhadap salah satu kelas gender. Nilai recall yang tinggi pada kelas Male dan Female menunjukkan bahwa model memiliki sensitivitas yang baik dalam mengenali masing-masing kelas. Sementara itu, nilai precision yang tinggi menandakan bahwa sebagian besar hasil prediksi yang diberikan oleh model merupakan prediksi yang benar. Konsistensi antara nilai precision dan recall pada kedua kelas menunjukkan bahwa model klasifikasi yang dihasilkan bersifat stabil dan dapat diandalkan. Secara keseluruhan, hasil Performance Vector membuktikan bahwa algoritma Decision Tree C4.5 dengan konfigurasi yang digunakan mampu menghasilkan model klasifikasi gender yang akurat, seimbang, dan mudah diinterpretasikan, sehingga layak digunakan sebagai pendekatan awal dalam sistem klasifikasi gender berbasis ciri fisik.



Gambar 8. Hasil *Decision Tree*

Hasil visualisasi struktur pohon keputusan menunjukkan bagaimana algoritma Decision Tree C4.5 melakukan proses pemilihan atribut secara bertingkat untuk mengklasifikasikan gender menjadi kelas Male dan Female berdasarkan ciri fisik. Atribut *nose_wide* muncul sebagai root node, yang menandakan bahwa lebar hidung merupakan atribut paling informatif dalam membedakan gender pada dataset Gender Classification v7. Pemilihan atribut ini didasarkan pada nilai information gain tertinggi yang diperoleh pada tahap awal pembentukan pohon.

Pada cabang selanjutnya, atribut *forehead_width_cm* dan *nose_long* digunakan sebagai pemisah tambahan, yang menunjukkan bahwa ukuran dahi dan panjang hidung memiliki kontribusi signifikan dalam proses klasifikasi. Nilai ambang batas (threshold) pada atribut numerik seperti *forehead_width_cm* dan *forehead_height_cm* membagi data ke dalam sub-kelas yang lebih homogen, sehingga meningkatkan ketepatan prediksi. Sementara itu, atribut kategorikal biner seperti *lips_thin* dan *distance_nose_to_lip_long* berperan dalam memperjelas keputusan pada kondisi data yang memiliki karakteristik fisik yang saling tumpang tindih.

Struktur pohon keputusan memperlihatkan bahwa sebagian besar jalur keputusan mengarah secara konsisten pada satu kelas dominan, baik Male maupun Female, dengan jumlah percabangan yang masih dapat dikendalikan. Hal ini menunjukkan bahwa model tidak mengalami kompleksitas berlebihan (overfitting) dan tetap mempertahankan kemampuan generalisasi yang baik terhadap data uji. Daun pohon (leaf node) yang dihasilkan merepresentasikan keputusan akhir klasifikasi gender berdasarkan kombinasi atribut fisik yang berbeda.

Secara keseluruhan, hasil Decision Tree membuktikan bahwa algoritma C4.5 mampu menghasilkan model klasifikasi yang mudah diinterpretasikan, dengan jalur keputusan yang jelas dan logis. Kombinasi antara atribut ukuran dahi, karakteristik hidung, serta ciri bibir menunjukkan bahwa ciri fisik memiliki peran penting dalam membedakan gender. Dengan struktur pohon yang relatif sederhana namun akurat, model ini layak digunakan sebagai pendekatan klasifikasi gender berbasis ciri fisik pada dataset Gender Classification v7.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Decision Tree C4.5 mampu diterapkan secara efektif untuk melakukan klasifikasi gender berdasarkan ciri fisik pada Dataset Gender Classification v7. Proses penelitian yang meliputi pengumpulan data, preprocessing, penentuan label, pembagian data latih dan data uji, pelatihan model, serta evaluasi kinerja menunjukkan bahwa dataset yang digunakan memiliki karakteristik yang sesuai untuk pendekatan klasifikasi berbasis pohon keputusan.

Hasil pengujian model menggunakan Altair AI Studio menghasilkan tingkat akurasi sebesar 97,20%, yang menunjukkan bahwa model mampu mengklasifikasikan data gender dengan tingkat ketepatan yang sangat baik. Nilai precision dan recall yang tinggi pada kedua kelas, yaitu Male dan Female, menandakan bahwa model memiliki kemampuan prediksi yang seimbang dan tidak bias terhadap salah satu kelas. Rendahnya tingkat kesalahan klasifikasi juga mengindikasikan bahwa atribut-atribut ciri fisik yang digunakan memiliki kontribusi signifikan dalam membedakan gender.

Visualisasi struktur pohon keputusan memperlihatkan bahwa atribut seperti lebar hidung, lebar dahi, tinggi dahi, serta karakteristik bibir dan jarak hidung ke bibir merupakan faktor penting dalam proses pengambilan keputusan. Struktur pohon yang dihasilkan relatif sederhana dan mudah diinterpretasikan, sehingga memberikan keunggulan dari sisi transparansi dan pemahaman model dibandingkan metode klasifikasi yang bersifat kompleks.

Secara keseluruhan, penelitian ini membuktikan bahwa algoritma Decision Tree C4.5 tidak hanya mampu menghasilkan model klasifikasi gender yang akurat, tetapi juga memberikan model yang interpretatif dan efisien. Oleh karena itu, metode ini layak digunakan sebagai pendekatan awal dalam pengembangan sistem klasifikasi gender berbasis ciri fisik dan dapat dikembangkan lebih lanjut dengan penambahan atribut atau perbandingan dengan algoritma klasifikasi lainnya untuk meningkatkan performa sistem.

DAFTAR PUSTAKA

- Algoritma, C., Carolina, A., Ade, K., & Kunci, K. (2020). *Penerapan Data Mining dengan Menggunakan Provinsi di Indonesia Pendahuluan*. 19, 27–38.
- Amaliyah, A., & Fatah, Z. (2024). *Implementasi Prediksi Penyakit Ginjal Kronis dengan Menggunakan Metode Decision Tree*. 3(2), 180–186.
- Dan, M., Pada, M., & Numerik, D. (2022). *Implementasi algoritma decision tree c4.5 dengan improvisasi mean dan median pada dataset numerik*. 5, 105–114. <https://doi.org/10.37600/tekinkom.v5i1.435>
- Kunci, K. (2023). *Indonesian Journal of Computer Science*. 12(1), 4228–4242.
- Podstawski, R., Boryślawski, K., Zsolt, B., & Alföldi, Z. (2022). *Sex Differences in Anthropometric and Physiological Profiles of Hungarian Rowers of Different Ages*.
- Science, C. (2022). *A REVIEW : CLASSIFICATION BASED DECISION TREE INDUCTION*. 20(19), 535–540. <https://doi.org/10.48047/nq.2022.20.19.NQ99049>
- Siregar, S. (2024). *JU-KOMI JU-KOMI*. 2(02), 30–37.
- Studi, P., Informatika, T., Teknik, F., Ponorogo, U. M., & Ponorogo, K. (n.d.). *Sistem Klasifikasi Penerima Bantuan Sosial Dengan Algoritma Decision Tree*. 7–16.
- Su, Y., Ho, C., Lee, P., Lin, C., Hung, Y., & Chen, P. (2023). *Gender and Age Differences in Anthropometric Characteristics of Taiwanese Older Adults Aged 65 Years and Older*. 1993, 1–16.