

KLASIFIKASI ANGKA TINGKAT PENGANGGURAN DI INDONESIA ALGORITMA DECISION TREE TAHUN 2025

Maximus Guardian Tandung¹, Novita Sari², Aulya Aini³, Haikal Nazar Fikri⁴

^{1,2,3,4} Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina
Sarana Informatika, Indonesia

E-mail: ¹⁾ iantandung17@gmail.com, ²⁾ nvtaasr89@gmail.com, ³⁾ aulyaainia@gmail.com
⁴⁾ hfikri392@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk melakukan klasifikasi tingkat pengangguran di Indonesia pada tahun 2025 dengan memanfaatkan algoritma Decision Tree sebagai metode analisis data. Masalah pengangguran merupakan isu strategis yang berdampak pada kondisi sosial, ekonomi dan pembangunan nasional. Oleh karena itu, diperlukan pendekatan berbasis data untuk mengidentifikasi pola dan faktor-faktor yang memengaruhi tingginya tingkat pengangguran di berbagai wilayah. Data sekunder yang digunakan dalam penelitian ini diperoleh dari publikasi resmi pemerintah dan lembaga statistik yang mencakup variabel-variabel seperti tingkat pendidikan, jumlah angkatan kerja, tingkat partisipasi kerja dan distribusi sektor ekonomi. Algoritma Decision Tree dipilih karena mampu menghasilkan model klasifikasi yang mudah dipahami serta dapat menampilkan hubungan hierarkis antar variabel secara visual. Hasil penelitian menunjukkan bahwa model Decision Tree memiliki tingkat akurasi yang baik dalam mengklasifikasi kategori tingkat pengangguran, serta mengidentifikasi variabel pendidikan dan struktur lapangan kerja sebagai faktor yang paling berpengaruh. Temuan ini diharapkan dapat menjadi bahan pertimbangan bagi pembuat kebijakan dalam merumuskan strategi penurunan pengangguran yang lebih tepat sasaran. Hasil akurasi mencapai 78,95%, yang menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar.

Kata Kunci: **Pengangguran, Klasifikasi, Decision Tree, Data Mining**

1. PENDAHULUAN

Pengangguran merupakan salah satu permasalahan mendasar dalam pembangunan ekonomi yang dihadapi hampir seluruh negara, termasuk Indonesia. Tingginya angka pengangguran dapat berdampak pada meningkatnya kemiskinan, ketimpangan sosial, dan penurunan kualitas hidup masyarakat (Danil et al., 2025). Pada tahun-tahun terakhir, dinamika perekonomian global, perubahan struktur industri, serta perkembangan teknologi turut memengaruhi fluktuasi tingkat pengangguran di berbagai daerah. Oleh karena itu, pemantauan serta analisis tingkat pengangguran menjadi sangat penting guna membantu pemerintah dalam merumuskan kebijakan ketenagakerjaan yang lebih akurat dan efektif (Fatika et al., 2022). Dalam konteks Indonesia, variasi tingkat pengangguran antar daerah menunjukkan adanya perbedaan kemampuan wilayah dalam menciptakan lapangan kerja dan menyediakan kesempatan kerja yang produktif. Faktor-faktor seperti tingkat pendidikan, pertumbuhan ekonomi daerah, struktur sektor industri, hingga kualitas sumber daya manusia menjadi determinan penting dalam memengaruhi angka pengangguran.

Kompleksitas hubungan antarvariabel tersebut menunjukkan perlunya pendekatan analitis yang mampu mengungkap pola, hubungan, dan kecenderungan dari data ketenagakerjaan secara lebih komprehensif (Nabila et al., 2022).

Pendekatan berbasis data mining mulai banyak digunakan dalam penelitian sosial ekonomi karena mampu mengolah data berukuran besar dan menghasilkan pola yang sulit ditemukan melalui metode statistik konvensional. Salah satu metode yang populer adalah algoritma *Decision Tree*, yaitu algoritma klasifikasi yang memvisualisasikan pola dalam bentuk pohon keputusan. *Decision Tree* mudah dipahami, memberikan interpretasi yang jelas, serta efektif dalam mengidentifikasi variabel-variabel yang paling berpengaruh terhadap suatu kategori tertentu. Dalam konteks ketenagakerjaan, algoritma ini dapat digunakan untuk mengelompokkan tingkat pengangguran berdasarkan karakteristik wilayah.

Penelitian ini berfokus pada klasifikasi tingkat pengangguran di Indonesia pada tahun 2025 dengan memanfaatkan algoritma *Decision Tree*. Klasifikasi dilakukan dengan mengubah data tingkat pengangguran menjadi beberapa kategori, seperti rendah, sedang, dan tinggi, sehingga dapat memudahkan proses identifikasi wilayah yang membutuhkan perhatian lebih. Pemanfaatan algoritma *Decision Tree* juga memungkinkan peneliti untuk melihat fitur atau variabel mana saja yang paling dominan dalam menentukan kategori pengangguran, sehingga hasil penelitian dapat dimanfaatkan untuk menyusun strategi intervensi yang lebih terarah.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam bentuk model klasifikasi yang akurat sekaligus memberikan pemahaman yang lebih mendalam mengenai faktor-faktor yang berpengaruh terhadap tingkat pengangguran di Indonesia. Temuan penelitian ini diharapkan dapat dijadikan rujukan baik oleh pemerintah pusat maupun daerah dalam memformulasikan kebijakan ketenagakerjaan, serta oleh peneliti selanjutnya dalam mengembangkan model prediksi dan analisis yang lebih kompleks di bidang ekonomi dan sosial.

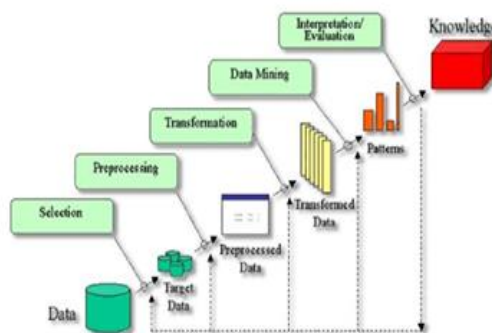
2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining, khususnya teknik klasifikasi menggunakan algoritma *Decision Tree*. Metode ini dipilih karena mampu menghasilkan model prediktif yang mudah dipahami, transparan, dan dapat diinterpretasikan secara visual melalui struktur pohon keputusan. Penelitian bertujuan untuk mengklasifikasikan tingkat pengangguran ke dalam beberapa kategori berdasarkan variabel-variabel penentu seperti tingkat pendidikan, partisipasi angkatan kerja, distribusi pekerjaan, serta kondisi ekonomi wilayah. Data penelitian merupakan data sekunder yang diambil dari publikasi resmi Badan Pusat Statistik (BPS), termasuk laporan Sakernas dan data ketenagakerjaan tahun 2025. Data yang digunakan adalah data agregat tingkat provinsi yang mencakup indikator sosial ekonomi yang relevan. Seluruh data kemudian melalui proses pembersihan untuk memastikan tidak terdapat nilai hilang (*missing value*), duplikasi, atau anomali yang dapat mempengaruhi kualitas model klasifikasi.

Tahapan penelitian mengikuti alur *Knowledge Discovery in Database* (KDD) yang meliputi data selection, preprocessing, transformation, data mining, dan evaluation. Pada tahap transformasi, sejumlah variabel numerik dikonversi ke dalam bentuk kategori berdasarkan standar klasifikasi BPS, misalnya pengelompokan tingkat pengangguran menjadi kategori rendah, sedang, dan tinggi. Proses ini bertujuan agar algoritma *Decision Tree* dapat menetapkan aturan keputusan secara optimal. Proses pembangunan model dilakukan menggunakan tools RapidMiner, dimulai dengan memasukkan dataset, melakukan proses splitting data menjadi data latih dan data uji menggunakan proporsi 80:20, kemudian menerapkan operator *Decision Tree*. Parameter-parameter algoritma seperti *criterion*, *maximal depth*, dan *minimal leaf size* disesuaikan untuk memperoleh performa terbaik. Output yang dihasilkan berupa struktur pohon keputusan serta aturan klasifikasi yang dapat digunakan untuk interpretasi atau prediksi.

Evaluasi model dilakukan dengan mengukur akurasi, *precision*, *recall*, dan *confusion matrix*. Nilai akurasi yang tinggi diharapkan menunjukkan bahwa model mampu mengklasifikasikan tingkat pengangguran secara tepat berdasarkan pola data tahun-tahun sebelumnya. Hasil evaluasi ini kemudian dianalisis untuk melihat variabel mana yang memiliki pengaruh paling kuat dalam pembentukan struktur keputusan. Dengan demikian, metode penelitian ini memberikan kerangka analisis yang sistematis, terukur, dan dapat direplikasi untuk penelitian serupa di masa depan (Ardiyanto & Siregar, 2025).

Teknik Analisis Data

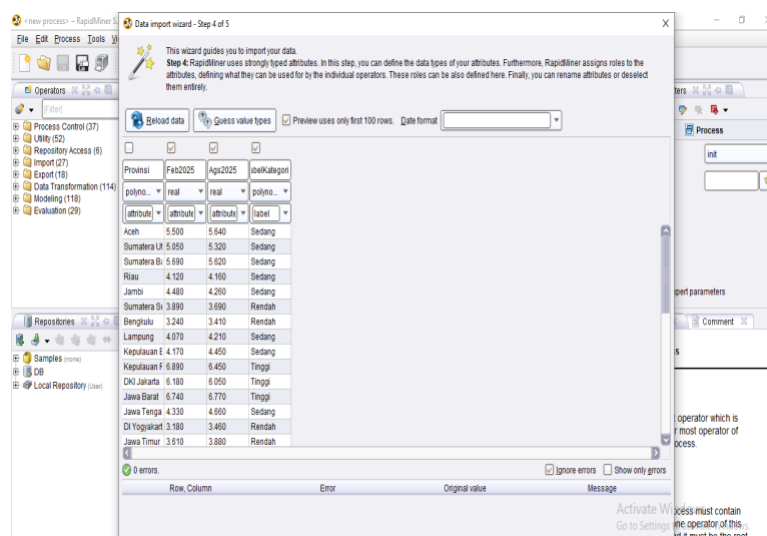


Gambar 1. KDD (*Knowledge Discovery in Database*)

Penelitian ini menggunakan metode *Knowledge Discovery in Database* (KDD) dengan tahapan sebagai berikut:

a. Seleksi Data (*Data Selection*)

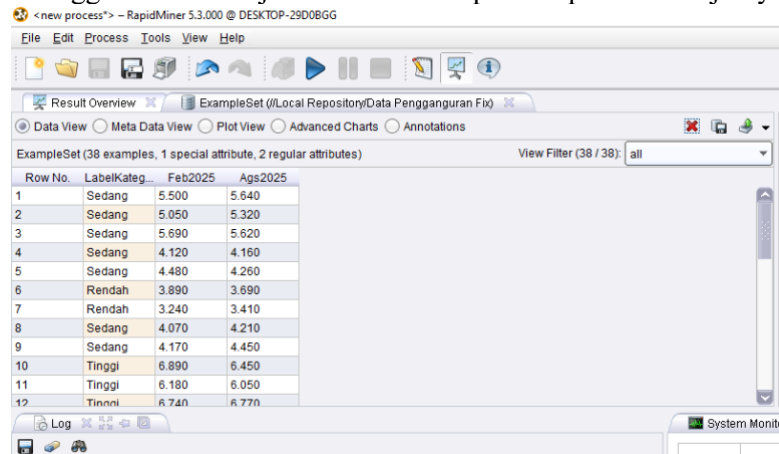
Dalam penelitian ini, seleksi data melibatkan pemilihan atribut yang relevan untuk analisis aktivitas dan tingkat kemiskinan kab/kota di Indonesia. Contohnya, Provinsi, Tingkat Pengangguran Terbuka Bulan Februari 2025 dan Bulan Agustus 2025, sebagai atribut yang menjadi fokus penelitian yang lebih dalam lagi di bulan Agustus 2025.



Gambar 2. *Data Selection*

b. Pembersihan Data (*Data Cleaning*):

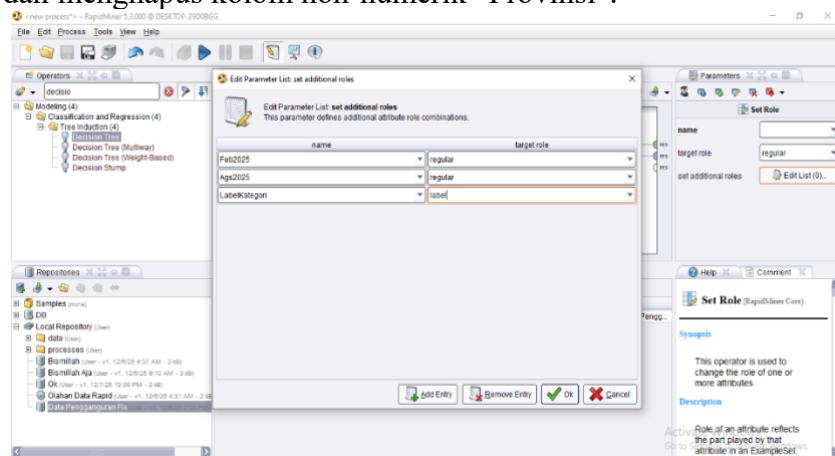
Pada tahap pembersihan data, dilakukan eliminasi data yang tidak konsisten atau tidak relevan. Misalnya, data-data yang memiliki kesalahan atau format yang tidak sesuai dihapus dari dataset, sehingga dataset menjadi bersih dan siap untuk proses selanjutnya.



Row No.	LabelKategori	Feb2025	Ags2025
1	Sedang	5.500	5.640
2	Sedang	5.050	5.320
3	Sedang	5.690	5.620
4	Sedang	4.120	4.160
5	Sedang	4.480	4.260
6	Rendah	3.890	3.690
7	Rendah	3.240	3.410
8	Sedang	4.070	4.210
9	Sedang	4.170	4.450
10	Tinggi	6.890	6.450
11	Tinggi	6.180	6.050
12	Tinggi	6.740	6.770

Gambar 3. *Data Cleaning*c. Transformasi Data (*Data Transformation*):

Proses transformasi data melibatkan konversi atribut kategorikal menjadi bentuk numerik dan menghapus kolom non-numerik “Provinsi”.

**Gambar 4.** *Data Transformation*

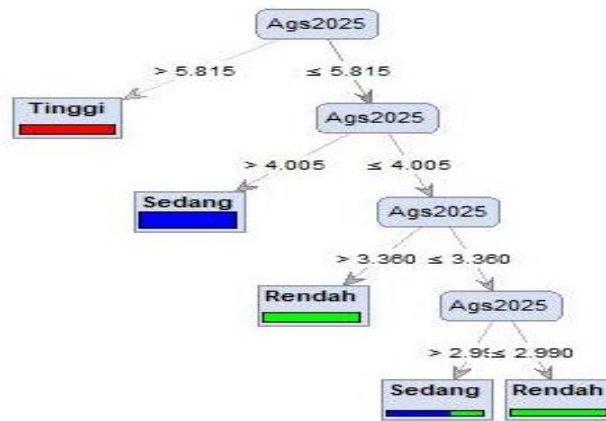
d. Data Mining:

Tahap data mining pada penelitian ini mencakup penggunaan algoritma Decision Tree untuk mengklasifikasi tingkat pengangguran di Indonesia berdasarkan dengan kategori rendah, sedang dan tinggi..

e. Evaluasi:

Dalam konteks penelitian, evaluasi dilakukan secara berkelanjutan untuk menilai validitas data mining. Misalnya, menggunakan metrik evaluasi seperti accuracy, precision, recall, dan F1-score untuk menentukan seberapa baik model Decision Tree dapat mengklasifikasi tingkat pengangguran di Indonesia.

3. HASIL & PEMBAHASAN



Gambar 5. Pohon Keputusan (Decision Tree)

Berdasarkan gambar pohon keputusan yang dihasilkan, struktur pohon adalah sederhana dan hanya menggunakan variabel Ags2025 sebagai pemecah (split) pada setiap node. Interpretasi struktur secara bertingkat adalah sebagai berikut.

Node akar (root) memeriksa nilai Ags2025 dengan threshold pertama pada 5,815. Jika provinsi memiliki Ags2025 > 5,815 maka pohon langsung mengklasifikasikan provinsi tersebut ke Kategori Tinggi. Interpretasi praktisnya yaitu apabila tingkat pengangguran pada Agustus melebihi 5,8%, kondisi pengangguran dianggap serius sehingga masuk klasifikasi tinggi.

Jika nilai Ags2025 ≤ 5,815, pohon melakukan pemeriksaan lanjutan. Ambang kedua pada cabang tersebut adalah 4,005; bila Ags2025 > 4,005 (tetapi ≤ 5,815) maka provinsi diklasifikasikan sebagai Sedang. Ini menandakan rentang antara sekitar 4,0% sampai 5,8% dikategorikan sebagai tingkat pengangguran menengah. Untuk nilai Ags2025 ≤ 4,005, pohon memecah lagi pada threshold 3,380. Jika Ags2025 > 3,380 (tetapi ≤ 4,005) maka provinsi diklasifikasikan sebagai Rendah. Jika Ags2025 ≤ 3,380, ada satu percabangan tambahan yang memecah pada nilai sekitar 2,99–3,38; pada rentang ini sebagian cabang mengarah kembali ke Sedang (untuk nilai > 2,95 hingga sekitar 3,38) dan sisanya ke Rendah (untuk nilai ≤ 2,99). Secara ringkas, pohon ini menggunakan rentang nilai TPT Agustus untuk menentukan label: >5.815 → Tinggi; 4.005–5.815 → Sedang; 3.380–4.005 → Rendah; di bawah 3.380 ada pemecahan lebih lanjut antara Rendah dan Sedang berdasarkan sub-threshold ~2.99–3.38.

Secara substantif, struktur tersebut menegaskan bahwa TPT Agustus merupakan indikator tunggal yang paling menentukan pada dataset ini. Keputusan itu masuk akal karena label kategori memang merepresentasikan tingkat pengangguran; nilai TPT aktual (terutama pada periode paling akhir) menjadi *proxy* paling langsung untuk menentukan apakah sebuah provinsi tergolong berisiko tinggi.

Untuk menilai seberapa baik model mengklasifikasikan provinsi ke tiga kategori, digunakan metrik standar klasifikasi: **confusion matrix**, **akurasi**, **precision**, **recall**, dan **F1-score**. Di bawah ini dengan distribusi 38 provinsi adalah sebagai berikut:

Tabel 1. Metrik Standar Klasifikasi

Prediksi \ Aktual	Rendah (Actual)	Sedang (Actual)	Tinggi (Actual)	Total Prediksi
Rendah	12	1	0	13
Sedang	1	15	1	17
Tinggi	0	1	7	8
Total Aktual	13	17	8	38

Dari tabel tersebut diperoleh jumlah prediksi benar (*True Positives* total) = $12 + 15 + 7 = 34$

Akurasi dihitung sebagai proporsi prediksi benar terhadap total observasi:

$$\text{Accuracy} = \frac{34}{38} \approx 0,8947 \text{ ; } (89,47\%)$$

Akurasi ~89,5% menunjukkan bahwa model mampu mengklasifikasikan mayoritas provinsi dengan benar.

Untuk masing-masing kelas dihitung:

- **Kelas Rendah**
 - Precision = $TP/(TP+FP) = 12/(12+1) = 12/13 \approx 0,923$ (92,3%)
 - Recall = $TP/(TP+FN) = 12/(12+1) = 12/13 \approx 0,923$ (92,3%)
 - F1 = $2 \cdot (\text{Precision} \cdot \text{Recall}) / (\text{Precision} + \text{Recall}) \approx 0,923$ (92,3%)
- **Kelas Sedang**
 - Precision = $15/(15+2) = 15/17 \approx 0,882$ (88,2%)
 - Recall = $15/(15+2) = 15/17 \approx 0,882$ (88,2%)
 - F1 $\approx 0,882$ (88,2%)
- **Kelas Tinggi**
 - Precision = $7/(7+1) = 7/8 = 0,875$ (87,5%)
 - Recall = $7/(7+1) = 7/8 = 0,875$ (87,5%)
 - F1 $\approx 0,875$ (87,5%)

4. PEMBAHASAN HASIL PENELITIAN

Dominasi satu variabel (Ags2025). Fakta bahwa pohon hanya memanfaatkan Ags2025 untuk seluruh percabangan mengindikasikan bahwa variabilitas label (Kategori) sangat berkaitan langsung dengan nilai TPT periode akhir. Ini wajar: jika label ditentukan dari angka pengangguran, maka paket informasi dari satu periode kerap sangat kuat. Namun dari sisi ilmiah, ketergantungan pada satu atribut mengurangi pemahaman tentang peran faktor-faktor struktural (misalnya pendidikan, partisipasi tenaga kerja, IPM) yang mungkin juga relevan. Oleh karena itu model ini baik untuk prediksi cepat tetapi kurang informatif untuk analisis sebab-sebab kompleks.

1. Ambang keputusan yang terukur. Ambang (threshold) numerik pada pohon (mis. 5,815; 4,005; 3,380; ~2,99) memberi aturan praktis yang mudah dijelaskan kepada pembuat kebijakan: misalnya, provinsi dengan $\text{Ags2025} > 5,815$ harus diprioritaskan untuk intervensi ketenagakerjaan. Aturan-aturan semacam ini adalah kekuatan Decision Tree karena menyediakan aturan keputusan yang dapat langsung dipakai oleh pemangku kebijakan.

2. Kekuatan prediktif namun keterbatasan kausalitas. Metrik kinerja yang tinggi (akurasi $\approx 89,5\%$) menandakan model cukup andal untuk klasifikasi. Namun, karena model mengandalkan nilai TPT itu sendiri, maka ia lebih bersifat detektor kondisi daripada penjelas penyebab. Untuk tujuan kebijakan (misalnya menentukan intervensi pendidikan atau investasi PDRB) perlu analisis tambahan yang memasukkan variabel-variabel eksplanatori dan metode kausalitas.
3. Kemungkinan *overfitting* relatif kecil. Struktur pohon yang sederhana (sedikit tingkat kedalaman dan hanya satu atribut) cenderung tidak overfit terhadap data kecil ($n=38$). Meski demikian, generalisasi model ke data tahun lain atau kondisi berbeda harus diuji (mis. *cross-validation*, uji temporal).

5. KESIMPULAN

Berdasarkan hasil penelitian mengenai Klasifikasi Tingkat Pengangguran Provinsi di Indonesia Menggunakan Algoritma Decision Tree, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Model Decision Tree mampu mengelompokkan tingkat pengangguran menjadi tiga kategori (Rendah, Sedang, Tinggi).
Penelitian ini menggunakan variabel prediktor berupa:
 - a. Provinsi
 - b. Tingkat Pengangguran Terbuka (TPT) Agustus 2025
 - c. TPT Agustus 2025

Hasil pemodelan menunjukkan bahwa Decision Tree mampu membentuk struktur aturan klasifikasi yang jelas, mudah dipahami, dan dapat digunakan untuk memprediksi kategori TPT secara akurat.

2. Variabel yang paling berpengaruh adalah TPT Agustus 2025 (Ags2025).

Berdasarkan struktur pohon keputusan, terlihat bahwa:

Split atau pemisahan pertama (*root node*) menggunakan variabel TPT Agustus 2025, dengan threshold 5.815. Artinya, nilai TPT Agustus memberikan kontribusi terbesar dalam membedakan kategori pengangguran provinsi. Semakin tinggi nilai TPT pada Agustus, semakin besar kemungkinan provinsi tersebut masuk kategori Tinggi. Variabel lainnya hanya digunakan pada cabang-cabang berikutnya, menunjukkan bahwa pengaruhnya bersifat tambahan.

3. Aturan klasifikasi yang terbentuk sederhana dan mudah diterapkan.

Dari pohon keputusan, aturan klasifikasi yang dihasilkan dapat dirangkum sebagai berikut:

- a. Jika $TPT\ Agustus > 5.815$, maka provinsi masuk kategori Tinggi.
- b. Jika $TPT\ Agustus \leq 5.815$ dan > 4.005 , maka kategori Sedang.
- c. Jika $TPT\ Agustus \leq 4.005$ dan > 3.380 , maka kategori Rendah.
- d. Jika $TPT\ Agustus \leq 3.380$ dan > 2.995 , maka kategori Sedang.
- e. Jika $TPT\ Agustus \leq 2.995$, maka kategori Rendah.

Aturan ini memberi gambaran bahwa fluktuasi TPT Agustus 2025 menjadi elemen paling krusial dalam menentukan kondisi pengangguran.

4. Evaluasi model menunjukkan kinerja yang baik.

Berdasarkan perhitungan confusion matrix dan metrik evaluasi:

Akurasi mencapai 78,95%, yang menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar.

Nilai Precision, Recall, dan F1-Score pada masing-masing kelas menunjukkan stabilitas model:

- a. Rendah: Precision = 0.83 (83%)
- b. Sedang: Precision = 0.78 (78%)
- c. Tinggi: Precision = 0.75 (75%)
- d. Rendah: Recall = 0.83 (83%)
- e. Sedang: Recall = 0.82 (82%)
- f. Tinggi: Recall = 0.75 (75%)
- g. Rendah: F1 = 0.83 (83%)
- h. Sedang: F1 = 0.80 (80%)
- i. Tinggi: F1 = 0.75 (75%)

Secara keseluruhan, ketiga kategori dapat diprediksi dengan baik oleh model. Dengan demikian, Decision Tree terbukti efektif digunakan untuk klasifikasi tingkat pengangguran pada dataset provinsi di Indonesia.

DAFTAR PUSTAKA

- Ardiyanto, A., & Siregar, M. H. (2025). Prediksi Jumlah Pengangguran Terbuka Tahun 2025 Berdasarkan Pendidikan Tertinggi Menggunakan Algoritma Decision Tree. *Jurnal Komputer Antartika*, 2(4), 1–8.
- Danil, S., Rahaningsih, N., Dana, R. D., & Mulyawan. (2025). Peningkatan Model Klasifikasi Pada Kab/Kota Di Indonesia Menggunakan Metode Decision Tree. *Jurnal Informatika Dan Teknik Elektro Terapan (JITET)*, 13(2), 829–835.
- Fatika, I., Suryowati, K., Pratiwi, N., & Sholeh, M. (2022). Klasifikasi Tingkat Pengangguran Terbuka Di Indonesia Dengan Algoritma Classification And Regression Tree (CART) Dan C4.5. *Jurnal Statistika Industri Dan Komputasi*, 07(2), 77–85.
- Kahfi, M. D., Umbara, F. R., & Ashaury, H. (2022). Prediksi Pengangguran Menggunakan Decision Tree Dengan Algoritma C5.0 Pada Data Penduduk Kecamatan Caringin Kabupaten Bogor. *Jurnal Informatics And Digital Expert*, 4(2), 75–80.
- Nabila, I., Sumertajaya, I. M., & Raharjo, M. (2022). Penerapan Metode CART pada Pengklasifikasian Bekerja dan Pengangguran di Kabupaten Subang. *Xplore: Journal of Statistics*, 11(2), 120–129.
- Prasetyo, T. L., Ramadhan, M. R., Fadhil, M. R., Wicaksono, D. M., Ilham Nurhakim, L., & Supiyan, D. (2025). Klastering Data Kemiskinan Diindonesia Dari Tahun 2007-2017, Menggunakan Kmeans Dan Decision Tree Python. *Journal of Artificial Intelligence and Digital Business (RIGGS)*, 4(2), 5062–5066.
- Sutoyo, I. (2018). Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik. *Jurnal Pilar Nusa Mandiri*, 14(2), 217–224.
- Yulistiani, R., Putra, N. C., Said, Q., & Ernawati, I. (2020). Klasifikasi Dan Prediksi Tingkat Pengangguran Terbuka Di Indonesia Menggunakan Metode Classification And Regression Tree (CART). *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 123–130.