

## PREDIKSI JUMLAH PERAWAT BERDASARKAN KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

Oleh:

**Aldi Patria Nugraha<sup>1</sup>**

**Tia Mulyani<sup>2</sup>**

**Elkin Rilvani<sup>3</sup>**

Universitas Pelita Bangsa

Alamat: Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Sel., Kabupaten Bekasi, Jawa Barat (17530).

Korespondensi Penulis: [patrianugrahaaldi20@gmail.com](mailto:patrianugrahaaldi20@gmail.com), [tiamulyani799@gmail.com](mailto:tiamulyani799@gmail.com), [elkin.rilvani@pelitabangsa.ac.id](mailto:elkin.rilvani@pelitabangsa.ac.id).

**Abstract.** The availability of healthcare personnel, particularly nurses, is a crucial aspect in ensuring equitable and high-quality healthcare services for all communities. West Java Province, as the most populous province in Indonesia, faces major challenges in the equitable distribution of nurses across various regencies and cities. This disparity can significantly affect the quality of healthcare services received by residents in certain regions, especially in remote and densely populated areas. This study aims to predict the number of nurses based on administrative regions using the Decision Tree algorithm. The research data were obtained from government agencies and include the number of nurses in each region over a specific period. The analysis process was carried out using RapidMiner software to build the predictive model and comprehensively evaluate its performance. The results indicate disparities in nurse distribution between regions, along with several variables that influence their distribution. This predictive model is expected to serve as a strategic basis for policymakers in formulating effective policies for the equitable allocation of healthcare personnel in West Java Province.

**Keywords:** Nurse, West Java, Decision Tree, Data Mining, RapidMiner.

---

Received July 10, 2025; Revised July 29, 2025; August 09, 2025

\*Corresponding author: [patrianugrahaaldi20@gmail.com](mailto:patrianugrahaaldi20@gmail.com)

# PREDIKSI JUMLAH PERAWAT BERDASARKAN KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

**Abstrak.** Ketersediaan tenaga kesehatan, khususnya perawat, merupakan aspek penting dalam menjamin layanan kesehatan yang merata dan berkualitas bagi seluruh masyarakat. Provinsi Jawa Barat, sebagai provinsi dengan jumlah penduduk terbanyak di Indonesia, menghadapi tantangan besar dalam pemerataan distribusi tenaga perawat di berbagai kabupaten dan kota. Ketimpangan ini dapat berdampak pada kualitas pelayanan kesehatan yang diterima oleh masyarakat di wilayah tertentu, terutama di daerah terpencil dan padat penduduk. Penelitian ini bertujuan untuk memprediksi jumlah perawat berdasarkan wilayah administratif menggunakan algoritma Decision Tree. Data penelitian diperoleh dari instansi pemerintah yang mencakup data jumlah perawat di setiap wilayah selama kurun waktu tertentu. Proses analisis dilakukan dengan bantuan perangkat lunak RapidMiner untuk membangun model prediksi dan mengevaluasi performanya secara menyeluruh. Hasil penelitian menunjukkan adanya disparitas jumlah perawat antarwilayah serta beberapa variabel yang memengaruhi distribusinya. Model prediksi ini diharapkan dapat menjadi dasar pengambilan keputusan strategis bagi pihak berwenang dalam merumuskan kebijakan pemerataan tenaga kesehatan di Provinsi Jawa Barat.

**Kata Kunci:** Perawat, Jawa Barat, Decision Tree, Data Mining, RapidMiner.

## LATAR BELAKANG

Perawat merupakan salah satu unsur vital dalam sistem pelayanan kesehatan, berperan sebagai ujung tombak dalam pelayanan kesehatan primer hingga perawatan lanjutan di rumah sakit. Ketersediaan dan pemerataan perawat sangat berpengaruh terhadap efektivitas sistem kesehatan secara menyeluruh, mulai dari puskesmas di daerah terpencil hingga rumah sakit rujukan di pusat kota. Namun, kenyataannya, distribusi perawat yang tidak merata masih menjadi salah satu permasalahan utama di Indonesia, termasuk di Provinsi Jawa Barat.

Provinsi Jawa Barat merupakan provinsi dengan jumlah penduduk terbanyak di Indonesia, sehingga memiliki kebutuhan tenaga kesehatan yang sangat besar dan kompleks. Beberapa kabupaten/kota mengalami kekurangan tenaga perawat, sedangkan daerah lain justru mengalami kelebihan. Ketimpangan ini tidak hanya berdampak pada kesenjangan mutu pelayanan kesehatan antar daerah, tetapi juga dapat memengaruhi

efisiensi anggaran, kepuasan pasien, dan beban kerja tenaga medis di fasilitas kesehatan tertentu.

Untuk itu, diperlukan pendekatan berbasis data yang mampu memberikan gambaran kondisi distribusi perawat secara objektif dan terukur. Salah satu teknik yang dapat digunakan adalah data mining, yaitu proses eksplorasi data untuk menemukan pola-pola tersembunyi yang dapat dijadikan dasar dalam pengambilan keputusan. Salah satu algoritma populer dalam data mining adalah *Decision Tree*, karena selain akurat dalam melakukan prediksi, metode ini juga menghasilkan struktur pohon keputusan yang mudah dipahami, bahkan oleh pengambil kebijakan non-teknis.

Penelitian ini bertujuan untuk menerapkan algoritma *Decision Tree* dalam membangun model prediksi jumlah perawat di setiap kabupaten/kota di Provinsi Jawa Barat. Dengan menggunakan data historis yang diperoleh dari instansi resmi, diharapkan model ini dapat mengidentifikasi faktor-faktor yang paling memengaruhi jumlah perawat dan memberikan wawasan yang berguna bagi Dinas Kesehatan Provinsi maupun pemerintah daerah dalam menyusun strategi pemerataan tenaga kesehatan secara lebih adil, efisien, dan berbasis bukti.

## KAJIAN TEORITIS

### Perawat (*Nurse*)

Perawat merupakan salah satu profesi di rumah sakit yang memiliki peran sentral dalam memberikan dan meningkatkan mutu pelayanan kesehatan. Untuk mewujudkan pelayanan kesehatan yang berkualitas, seorang perawat dituntut untuk memiliki profesionalisme yang tinggi.(Suprapto & Mulat, 2021). Sedangkan menurut Undang-Undang Republik Indonesia No. 44 Tahun 2019, perawat adalah individu yang telah menyelesaikan pendidikan tinggi keperawatan, baik di dalam maupun luar negeri, dan diakui secara resmi oleh pemerintah sesuai ketentuan yang berlaku. Sementara itu, berdasarkan International Labour Organization dalam *International Hazard Datasheets on Occupation (HDO)*, perawat didefinisikan sebagai tenaga kesehatan profesional yang terdaftar, yang bertugas membantu dokter dalam pelaksanaan tugas medis serta memberikan layanan keperawatan kepada pasien yang mengalami sakit, cedera, gangguan fisik maupun mental, serta kebutuhan kesehatan lainnya. (Dame Maria Pakpahan et al., 2023). Di Provinsi Jawa Barat, perawat tersebar di seluruh

# **PREDIKSI JUMLAH PERAWAT BERDASARKAN KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE***

kabupaten/kota, namun dengan jumlah dan distribusi yang belum merata. Ketimpangan ini dapat berdampak pada kesenjangan mutu pelayanan kesehatan antar wilayah, terutama pada daerah dengan keterbatasan sumber daya manusia kesehatan. Oleh karena itu, analisis terhadap distribusi dan prediksi jumlah perawat menjadi penting sebagai dasar perencanaan dan pengambilan kebijakan, agar alokasi tenaga keperawatan dapat lebih optimal dan sesuai dengan kebutuhan masing-masing daerah.

## **Data Mining**

Data mining merupakan istilah yang merujuk pada proses menemukan pengetahuan tersembunyi di dalam suatu basis data. Proses ini melibatkan penggunaan berbagai teknik seperti statistika, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi serta mengidentifikasi informasi yang relevan dan berguna dari kumpulan data dalam skala besar.

Data mining, yang juga dikenal sebagai *pattern recognition*, merupakan metode dalam pengolahan data yang bertujuan untuk menemukan pola-pola tersembunyi dari data yang telah dikumpulkan. Pola-pola tersebut kemudian diolah menjadi pengetahuan baru yang berasal dari data historis, dan hasilnya dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan di masa mendatang.

Tahapan proses dalam penerapan Data Mining merupakan bagian dari rangkaian *Knowledge Discovery in Database* (KDD), yang berkaitan dengan teknik integrasi, penemuan ilmiah, interpretasi, serta visualisasi pola-pola dari sejumlah data. Berikut ini disajikan ilustrasi mengenai tahapan-tahapan dalam proses data mining :

1. Data *cleaning*, yaitu proses untuk menghilangkan data yang tidak konsisten atau mengandung noise, dan data *integration*, yaitu tahap di mana berbagai sumber data yang terpisah digabungkan menjadi satu kesatuan yang utuh.
2. Data *selection* merupakan tahap di mana data yang dianggap relevan dengan tujuan analisis dipilih dan diambil kembali dari database untuk digunakan dalam proses selanjutnya.
3. Data *transformation* adalah proses mengubah atau menggabungkan data ke dalam format yang sesuai untuk proses penambangan, sering kali

melibatkan pembuatan ringkasan data atau penerapan operasi agregasi guna meningkatkan efisiensi analisis.

4. Data mining merupakan tahap utama di mana metode-metode cerdas digunakan untuk mengekstraksi pola atau informasi penting dari kumpulan data yang tersedia.
5. Pattern *evaluation* adalah proses untuk mengidentifikasi pola-pola yang benar-benar bermakna dan mewakili pengetahuan, berdasarkan pada sejumlah kriteria atau tindakan yang dianggap menarik dan relevan dalam konteks analisis.
6. *Knowledge presentation* merupakan tahap di mana teknik visualisasi dan penyajian pengetahuan digunakan untuk menyampaikan hasil temuan data mining secara jelas dan informatif kepada pengguna, sehingga dapat dengan mudah dipahami dan dimanfaatkan. (Riany & Testiana, 2023)

### **Algoritma *Decision Tree***

*Decision Tree* merupakan salah satu algoritma yang populer dan efektif dalam melakukan klasifikasi serta prediksi. Algoritma ini mampu merepresentasikan berbagai fakta ke dalam bentuk struktur pohon keputusan. Pohon keputusan sendiri adalah struktur yang digunakan untuk membagi kumpulan data besar menjadi bagian-bagian yang lebih kecil dan terstruktur. Atribut kelas berfungsi sebagai representasi untuk simpul daun (*leaf node*) dalam pohon keputusan. Sementara itu, simpul lainnya terdiri dari node internal yang terbentuk berdasarkan kondisi pengujian terhadap atribut pada sejumlah data dengan karakteristik tertentu, serta node akar (*root node*) yang menjadi titik awal dari struktur pohon.(Sephya et al., 2023). Decision Tree tepat digunakan untuk memprediksi jumlah perawat berdasarkan kabupaten/kota di Provinsi Jawa Barat karena dapat mengolah data kategorikal dan numerik secara bersamaan. Hal ini memungkinkan data wilayah kabupaten/kota dipadukan dengan variabel numerik seperti jumlah penduduk, jumlah rumah sakit, atau ketersediaan fasilitas kesehatan tanpa memerlukan proses transformasi data yang kompleks. Selain itu, model ini mampu menampilkan hasil dalam bentuk pohon keputusan yang mudah dipahami, sehingga memudahkan analisis faktor-faktor yang memengaruhi jumlah perawat di setiap wilayah.

# **PREDIKSI JUMLAH PERAWAT BERDASARKAN KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE***

## **Penggunaan RapidMiner dalam Data Mining**

RapidMiner adalah perangkat lunak berbasis *open source* yang dapat diakses secara bebas oleh siapa saja. Aplikasi ini digunakan sebagai solusi dalam melakukan analisis dan pemrosesan data. RapidMiner mendukung berbagai teknik analisis, termasuk teknik deskriptif dan prediktif. Dalam operasionalnya, RapidMiner dibangun dan dijalankan menggunakan bahasa pemrograman Java. (Sari et al., 2020). Berkat kemajuan teknologi algoritma komputasi dan analisis data berbasis komputer, proses pengolahan data dapat dilakukan dengan bantuan software RapidMiner. Algoritma *Decision Tree* akan digunakan dalam penelitian ini untuk memprediksi jumlah perawat berdasarkan data kabupaten/kota di Provinsi Jawa Barat. Hasil dari penelitian ini diharapkan dapat membantu pemerintah daerah dalam mengetahui distribusi tenaga perawat di setiap wilayah secara lebih akurat. Dengan demikian, pemerintah dapat mengambil kebijakan yang tepat dan cepat dalam pemerataan tenaga kesehatan, khususnya untuk meningkatkan kualitas pelayanan kesehatan di seluruh wilayah Provinsi Jawa Barat.

## **METODE PENELITIAN**

### **Metode Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk melakukan prediksi jumlah perawat di Provinsi Jawa Barat berdasarkan variabel wilayah kabupaten/kota dan faktor pendukung lainnya seperti jumlah penduduk dan fasilitas kesehatan. Metode ini dipilih karena mampu mengolah data historis dalam jumlah besar serta mengidentifikasi pola-pola tersembunyi yang dapat mendukung proses pengambilan keputusan. Algoritma *Decision Tree* digunakan sebagai teknik utama karena kemampuannya dalam membangun model prediktif yang mudah dipahami, dapat divisualisasikan secara hierarkis, serta mampu menunjukkan faktor-faktor yang paling berpengaruh terhadap hasil prediksi.

### **Metode Pengumpulan Data**

Pengumpulan data adalah tahap awal yang dilakukan untuk memperoleh data yang selanjutnya akan digunakan dalam proses pengolahan dan analisis. (Widodo & Setyawan, n.d.) .Data dalam penelitian ini berasal dari data sekunder yang diambil melalui portal resmi Open Data Provinsi Jawa Barat ([opendata.jabarprov.go.id](http://opendata.jabarprov.go.id)). Dataset mencakup:

- a. Tahun: 2019–2024
- b. Wilayah: Kabupaten/Kota di Provinsi Jawa Barat
- c. Jumlah perawat per wilayah

Data tersebut diperoleh dalam format .XLSX dan diolah menggunakan RapidMiner Studio, yaitu platform *visual data science* yang memungkinkan penerapan *machine learning* tanpa memerlukan pemrograman yang rumit.

## Tahapan Penelitian

Penelitian ini dilaksanakan secara sistematis melalui pendekatan data mining, dimulai dari tahap pengumpulan dan praprosesan data, pembuatan model prediksi dengan algoritma *Decision Tree*, hingga tahap evaluasi kinerja model. Seluruh rangkaian proses tersebut dilakukan menggunakan aplikasi RapidMiner, yang memungkinkan integrasi alur kerja secara visual dan efisien.

### 1. Praproses Data

Praproses data adalah tahapan yang bertujuan untuk mengubah dokumen teks ke dalam format kata yang lebih terstruktur dan jelas, dengan cara mereduksi atau menyederhanakan fitur-fitur pada data masukan agar data tersebut siap untuk dianalisis lebih lanjut.(Harahap & Muslim, 2020)

Adapun langkah-langkah praproses yang dilakukan meliputi:

- a) Pemeriksaan Konsistensi Data: Dataset ditinjau untuk memastikan tidak terdapat nilai kosong (*missing value*) maupun ketidaksesuaian data, seperti penulisan nama wilayah yang tidak seragam.
- b) Pembersihan Data: Menghilangkan data duplikat, baris kosong, serta entri yang mengandung karakter atau simbol yang tidak sesuai dengan standar input.
- c) Transformasi Atribut: Menyesuaikan nama kolom dan format penulisan data agar selaras dengan skema RapidMiner, termasuk mengonversi teks menjadi kategori atau nilai numerik (*label encoding*) jika diperlukan.

# PREDIKSI JUMLAH PERAWAT BERDASARKAN KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

- d) Seleksi Atribut: Atribut seperti "nama\_kabupaten\_kota" dan "tahun" ditetapkan sebagai tipe nominal, sedangkan "jumlah\_perawat" ditentukan sebagai label (target) dengan tipe numerik.
- e) Split Data: Dataset dibagi menjadi 70% untuk data pelatihan (*training*) dan 30% untuk data pengujian (*testing*) model, menggunakan metode pengambilan sampel acak (*random sampling*).

Langkah ini menjamin bahwa data yang digunakan telah bersih dan siap diolah, sehingga tidak menimbulkan bias dalam proses pembelajaran algoritma.

## 2. Pembangunan Model Decision Tree

Setelah data dipersiapkan, tahap berikutnya adalah membangun model klasifikasi menggunakan algoritma Decision Tree. Algoritma ini dipilih karena kemampuannya mengklasifikasikan data sekaligus menghasilkan model yang divisualisasikan dalam bentuk pohon keputusan, sehingga mudah dipahami, termasuk oleh pemangku kebijakan yang tidak memiliki latar belakang teknis.

Proses pembangunan model dilakukan melalui tahapan:

- a. Mengintegrasikan blok *Retrieve* dengan *Decision Tree* di dalam RapidMiner.
- b. Menyesuaikan parameter seperti *tree depth* (kedalaman pohon), *criterion* (misalnya *least square*), serta metode pruning untuk mencegah terjadinya *overfitting*.
- c. Melakukan pelatihan model dengan data training serta meninjau struktur pohon keputusan yang dihasilkan, termasuk node dan leaf yang merepresentasikan hubungan antar atribut.
- d. Mengidentifikasi variabel utama yang paling berpengaruh terhadap jumlah perawat berdasarkan hasil *feature selection* dari struktur pohon keputusan.

Model ini memungkinkan pengguna memahami aturan-aturan prediktif, seperti “Jika wilayah = Kabupaten Bogor dan tahun = 2019, maka jumlah perawat cenderung tinggi.”

### 3. Evaluasi Kinerja Model

Tahap terakhir dari proses penelitian ini adalah melakukan evaluasi terhadap performa model yang telah dibangun menggunakan data uji. Evaluasi ini bertujuan untuk menilai sejauh mana model mampu memprediksi jumlah perawat secara akurat dan andal. Proses pengujian dilakukan dengan memanfaatkan blok *Apply Model* dan *Performance* pada RapidMiner. Metrik evaluasi *Decision Tree* untuk regresi yang digunakan meliputi:

- a. *RMSE (Root Mean Squared Error)*: Menghitung rata-rata besar kesalahan prediksi yang dihasilkan model.
- b. *Absolute Error*: Menilai rata-rata selisih absolut antara nilai prediksi dan nilai aktual.
- c. *Relative Error*: Berfungsi untuk mengukur tingkat kesalahan relatif antara hasil prediksi dan nilai aktual.
- d. *Correlation*: Mengukur tingkat hubungan atau keterkaitan antara nilai prediksi dengan nilai aktual.

## HASIL DAN PEMBAHASAN

### Pengumpulan Data Dan Analisis

#### 1. Pengumpulan Data

Sebelum digunakan dan diolah, data tersebut terlebih dahulu melalui tahap persiapan (*preprocessing*) agar memiliki kelas binomial dan polinomial sesuai aturan yang telah ditetapkan serta berdasarkan nilai datanya. Hasil *preprocessing* sesuai dengan jenis kelas masing-masing data dapat dilihat pada tabel berikut :

***Tabel 4.1 Data Preprocessing***

# PREDIKSI JUMLAH PERAWAT BERDASARKAN KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

kode_provinil_id	nama_provinil_id	kode_kabupaten_id	nama_kabupaten_id	jumlah_perawat_id	satuan_perawat_id	#	tahun_id	##
1	JAWA BARAT	3201	KABUPATEN CIREBES	1011	1044KX	2018		
2	JAWA BARAT	3202	KABUPATEN SUMEDANG	1751	1044KJ	2018		
3	JAWA BARAT	3203	KABUPATEN CIMAHLA	1181	DRWNO	2018		
4	JAWA BARAT	3204	KABUPATEN BEKASI	1210	DRWNO	2018		
5	JAWA BARAT	3205	KABUPATEN GARUT	1910	DRWNO	2018		
6	JAWA BARAT	3206	KABUPATEN TAGAL	879	DRWNO	2018		
7	JAWA BARAT	3207	KABUPATEN CIREBES	1341	DRWNO	2018		
8	JAWA BARAT	3208	KABUPATEN KEBUMEN	1351	DRWNO	2018		
9	JAWA BARAT	3209	KABUPATEN CIREBES	1362	DRWNO	2018		
10	JAWA BARAT	3210	KABUPATEN SUMEDANG	1371	DRWNO	2018		
11	JAWA BARAT	3211	KABUPATEN SUMEDANG	1381	1044KX	2018		
12	JAWA BARAT	3212	KABUPATEN SUMEDANG	1391	1044KX	2018		
13	JAWA BARAT	3213	KABUPATEN SUMEDANG	1401	1044KX	2018		
14	JAWA BARAT	3214	KABUPATEN PURWAKARTA	1411	1044KX	2018		
15	JAWA BARAT	3215	KABUPATEN KARAWANG	2371	DRWNO	2018		
16	JAWA BARAT	3216	KABUPATEN BEKASI	1250	DRWNO	2018		
17	JAWA BARAT	3217	KABUPATEN BEKASI	891	DRWNO	2018		
18	JAWA BARAT	3218	KABUPATEN PANGKAL	318	DRWNO	2018		
19	JAWA BARAT	3219	6731A PT30190	2041	1044KX	2019		

## 2. Analisis Algoritma *Decission Tree*

Pada penelitian ini dilakukan prediksi dengan menggunakan algoritma decision tree pada RapidMiner dengan total 113 baris data sampel. Adapun langkah-langkah prediksi data tersebut adalah sebagai berikut :

- Siapkan data kemudian buka aplikasi RapidMiner, impor data menggunakan Read Excel, lalu pastikan kolom "nama\_kabupaten\_kota" dan "tahun" diatur sebagai tipe nominal. Selanjutnya, tetapkan kolom "jumlah\_perawat" sebagai label (target) dengan tipe numerik.
- Setelah data siap, buat proses *Decision Tree* dengan menambahkan operator *Set Role* untuk menetapkan "jumlah\_perawat" sebagai label. Kemudian, tambahkan operator *Decision Tree* dan hubungkan *output* dari *Set Role* ke *Decision Tree*. Selanjutnya, tambahkan operator *Split Data* untuk membagi dataset menjadi data *training* (70%) dan data *testing* (30%).
- Konfigurasikan parameter pada *Decision Tree* dengan memilih *least square* pada bagian *criterion*. Batasi nilai *maximum depth* hingga 10 untuk mencegah terjadinya *overfitting* pada model.
- Selanjutnya, jalankan serta evaluasi model dengan menambahkan operator *Apply Model* dan *Performance Regression*. Setelah seluruh

operator terhubung, jalankan proses, dan sistem akan menampilkan struktur pohon keputusan beserta nilai hasil prediksi.

## Evaluasi Dan Validasi Hasil

Penelitian ini mengevaluasi model menggunakan metrik RMSE, *Absolute Error*, *Relative Error*, dan *Correlation*. Proses untuk memperoleh nilai-nilai tersebut dapat dilihat pada gambar berikut:



Gambar 4.1 Process Design

Berikut adalah hasil tree yang dibuat dengan model KKLLayout :



Gambar 4.2 Hasil Regression Tree

Berikut merupakan hasil *performance vector* yang diperoleh dari algoritma *Decision Tree* yang digunakan dalam penelitian ini, yaitu:

# PREDIKSI JUMLAH PERAWAT BERDASARKAN KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

## PerformanceVector

```
PerformanceVector:  
root_mean_squared_error: 541.762 +/- 0.000  
absolute_error: 311.998 +/- 442.904  
relative_error: 26.89% +/- 48.67%  
squared_error: 293506.590 +/- 907442.182  
correlation: 0.927
```

Gambar 4.3 Hasil Performance Vector

Berdasarkan hasil yang diperoleh dari performance vector tersebut, dapat dijelaskan sebagai berikut:

1. *Root Mean Squared Error (RMSE)* =  $541.762 \pm 0.000$ .

Nilai ini menunjukkan rata-rata besar kesalahan prediksi model terhadap data aktual adalah sekitar 541 orang perawat. Semakin kecil nilai RMSE, semakin baik akurasi model.

2. *Absolute Error* =  $311.998 \pm 442.904$ .

Rata-rata selisih absolut antara nilai prediksi dan nilai sebenarnya adalah sekitar 312 orang perawat, dengan variasi kesalahan yang cukup besar ( $\pm 442,904$ ).

3. *Relative Error* =  $26.09\% \pm 48.67\%$ .

Kesalahan prediksi relatif terhadap nilai aktual berada pada kisaran 26,09% rata-rata, dengan tingkat variasi hingga  $\pm 48,67\%$ .

4. *Squared Error* =  $293.506,590 \pm 907.442,182$ .

Merupakan total kuadrat selisih antara nilai prediksi dan nilai aktual. Nilai ini lebih berguna dalam konteks perbandingan antar model, di mana nilai lebih rendah menunjukkan performa yang lebih baik.

5. *Correlation* = 0.927.

Koefisien korelasi antara hasil prediksi dan data aktual mencapai 0,927, yang mengindikasikan adanya hubungan yang sangat kuat dan positif. Nilai ini mendekati 1, yang berarti model memiliki kemampuan prediksi yang cukup baik dalam mengikuti pola data sebenarnya.

## KESIMPULAN DAN SARAN

### Kesimpulan

Penelitian ini menunjukkan bahwa algoritma *Decision Tree* mampu memodelkan dan memprediksi jumlah perawat di Kabupaten/Kota Provinsi Jawa Barat dengan performa yang cukup baik. Terbukti dari koefisien korelasi tinggi (0,927) yang mengindikasikan hubungan kuat antara prediksi dan nilai aktual. Namun demikian, terdapat kesalahan prediksi yang masih signifikan: RMSE sebesar 541,762 dan rata-rata absolute error sekitar 312 perawat, dengan relative error rata-rata ~26%, serta varians kesalahan yang tinggi, menunjukkan bahwa model kadang menyimpang cukup jauh pada beberapa observasi. Secara keseluruhan, model dapat dijadikan alat bantu prediktif, namun perlu kehati-hatian karena tidak sepenuhnya bebas dari ketidakpastian.

### Saran

1. Peningkatan Kualitas dan Kekayaan Fitur: Tambahkan variabel pendukung lain yang berpotensi memengaruhi distribusi perawat, seperti jumlah fasilitas kesehatan, tingkat pendidikan tenaga kesehatan, kepadatan penduduk, rasio urban-rural, dan indikator ekonomi wilayah, untuk menangkap variasi yang belum tertangkap model saat ini.
2. *Feature Engineering & Normalisasi*: Lakukan rekayasa fitur (misalnya rasio perawat per populasi, transformasi log untuk distribusi skewed) serta standardisasi/normalisasi bila perlu untuk membantu stabilitas dan akurasi pemodelan.
3. Validasi Silang dan Tuning: Terapkan cross-validation untuk mendapatkan estimasi performa yang lebih andal dan lakukan optimasi *hyperparameter* (kedalaman pohon, minimal *samples per leaf*, *pruning*) agar *overfitting/underfitting* dapat diminimalkan.
4. Metode *Ensemble*: Pertimbangkan penggunaan metode *ensemble* seperti *Random Forest* atau *Gradient Boosting* untuk mengurangi varians dan meningkatkan akurasi dibandingkan satu pohon keputusan tunggal.
5. Pembaruan Data Berkala: Perbarui dataset secara berkala dan verifikasi konsistensi agar model tetap relevan dengan dinamika perubahan jumlah perawat dan karakteristik wilayah.

# **PREDIKSI JUMLAH PERAWAT BERDASARKAN KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN DECISION TREE**

6. Evaluasi Kesalahan Ekstrem: Telaah kasus-kasus dengan error besar (*outlier*) untuk memahami apakah penyebabnya karena data, kondisi khusus wilayah, atau keterbatasan model, lalu tangani sesuai temuan.
7. Penggunaan Kebijakan: Gunakan output model sebagai alat bantu dalam pengambilan kebijakan distribusi tenaga perawat, bukan sebagai satu-satunya dasar. Kombinasikan dengan analisis kualitatif dan masukan lokal untuk keputusan yang lebih tepat sasaran.

## **DAFTAR REFERENSI**

- Dame Maria Pakpahan, Fitriany Suangga, & Rizki Sari Utami. (2023). Hubungan Karakteristik Perawat Dan Beban Kerja Dengan Kelelahan Kerja Perawat Di Ruang Rawat Inap RSUD Kota Tanjungpinang. *Jurnal Rumpun Ilmu Kesehatan*, 4(1), 10–27. <https://doi.org/10.55606/jrik.v4i1.2751>
- Harahap, R. N., & Muslim, K. (2020). Peningkatan Akurasi pada Prediksi Kepribadian Mbti Pengguna Twitter Menggunakan Augmentasi Data. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(4), 815. <https://doi.org/10.25126/jtiik.2020743622>
- Riany, A. F., & Testiana, G. (2023). Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes. *Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit ISPA Sonia*, 13(1), 42–54. <https://doi.org/10.33020/saintekom.v13i1.352>
- Sari, Y. R., Sudewa, A., Lestari, D. A., & Jaya, T. I. (2020). Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 192. <https://doi.org/10.24114/cess.v5i2.18519>
- Septya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, R., Irawan, Y., & Hayami, R. (2023). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 15–19. <https://doi.org/10.57152/malcom.v3i1.591>

- Suprapto, S., & Mulat, T. C. (2021). Faktor Determinan Pengembangan Kapasitas Perawat dalam Pelayanan Kesehatan. *Jurnal Ilmiah Kesehatan Sandi Husada*, 10(2), 416–422. <https://doi.org/10.35816/jiskh.v10i2.628>
- Widodo, F. T., & Setyawan, M. B. (n.d.). *Implementasi Text Mining Pada Analisis Sentimen Pemain Naturalisasi Timnas Indonesia Dengan Algoritma Naïve Bayes dan Support Vector Machine (SVM) 1 Pendahuluan.*