

Analisis Komparatif Support Vector Regression dan Decision Tree Regression untuk Peramalan Kasus HIV Jawa Barat

Risqi Agung Alamsyah¹, Safitri Juanita^{2*}

^{1,2}Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta

*safitri.juanita@budiluhur.ac.id

Abstrak

Peningkatan kasus HIV di Indonesia, termasuk di Provinsi Jawa Barat yang berada di antara wilayah dengan angka infeksi tertinggi, menekankan pentingnya ketersediaan model peramalan yang mampu memberikan estimasi akurat sebagai dasar perencanaan kebijakan kesehatan. Kondisi ini menuntut adanya pendekatan analitik yang dapat menangkap dinamika perkembangan kasus secara lebih komprehensif. Penelitian ini bertujuan mengembangkan model peramalan jumlah kasus HIV di Jawa Barat dengan menerapkan pendekatan CRISP-DM menggunakan dataset berdasarkan kelompok umur untuk periode 2019–2023. Dua algoritma regresi, yaitu Support Vector Regression (SVR) dan Decision Tree Regression (DTR), dibandingkan melalui tiga metrik evaluasi: Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), dan Mean Absolute Percentage Error (MAPE). Hasil penelitian menunjukkan bahwa DTR memberikan performa superior dengan R^2 sebesar 83.62, RMSE sebesar 32.29, dan MAPE sebesar 33.30. Sementara itu, SVR menghasilkan R^2 sebesar 36.33, RMSE sebesar 84.01, dan MAPE sebesar 44.87. Berdasarkan temuan tersebut, Decision Tree Regression diidentifikasi sebagai model yang lebih efektif untuk peramalan jumlah kasus HIV di Jawa Barat, yang dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam perencanaan program pencegahan, alokasi sumber daya kesehatan, serta penentuan prioritas intervensi di tingkat wilayah.

Kata kunci : Decision Tree Regression, HIV, Jawa Barat, Peramalan, Support Vector Regression

Abstract

The increasing number of HIV cases in Indonesia, particularly in West Java, which ranks among the regions with the highest infection rates, highlights the need for forecasting models capable of producing accurate estimates to support health policy planning. This situation underscores the importance of analytical approaches that can capture the dynamic progression of cases over time. This study aims to develop a forecasting model for the number of HIV cases in West Java by applying the CRISP-DM framework and utilizing a dataset categorized by age group for the period 2019–2023. Two regression algorithms, Support Vector Regression (SVR) and Decision Tree Regression (DTR), were compared using three evaluation metrics: Coefficient of Determination (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The findings indicate that DTR delivers superior performance, achieving an R^2 of 83.62, RMSE of 32.29, and MAPE of 33.30. In contrast, SVR produced an R^2 of 36.33, RMSE of 84.01, and MAPE of 44.87. Based on these results, Decision Tree Regression is identified as the more effective model for forecasting the number of HIV cases in West Java, providing practical support for health policy decision-making, resource allocation, and prioritization of targeted prevention and intervention programs at the regional level..

Keywords : Decision Tree Regression, Forecasting, HIV, Support Vector Regression, West Java

1. Pendahuluan

HIV (*Human Immunodeficiency Virus*) merupakan salah satu isu kesehatan global yang terus

menjadi perhatian karena karakteristik penularannya dan dampaknya terhadap kualitas hidup masyarakat. Di Indonesia, persebaran kasus HIV menunjukkan tren yang bervariasi

antarprovinsi, dan Jawa Barat merupakan salah satu wilayah dengan angka kasus yang cukup tinggi [1]. Data tahun 2022 mencatat bahwa Jawa Barat menempati peringkat kedua secara nasional dengan 1.115 orang yang hidup dengan HIV/AIDS (ODHA) di wilayah tersebut [2]. Kondisi ini menegaskan perlunya upaya prediktif untuk memahami potensi peningkatan kasus di masa depan, sehingga strategi penanggulangan dan alokasi sumber daya kesehatan dapat direncanakan secara lebih tepat dan efektif.

Berdasarkan penelitian tahun 2024, faktor-faktor yang mempengaruhi kejadian HIV/AIDS di Jawa Barat, antara lain adalah jumlah kasus kekerasan, persentase penggunaan alat kontrasepsi kondom pada anggota program Keluarga Berencana (KB), kepadatan penduduk, persentase penduduk miskin, dan jumlah fasilitas kesehatan per 100.000 penduduk [3].

Data penyebaran HIV di Provinsi Jawa Barat telah tersedia secara terbuka melalui portal *Open Data Jawa Barat* dalam bentuk jumlah kasus HIV pada periode 2019–2023. Namun demikian, ketersediaan data tersebut hingga saat ini masih bersifat deskriptif dan belum dimanfaatkan secara optimal sebagai sumber informasi prediktif yang dapat mendukung perencanaan dan pengambilan keputusan strategis oleh pemerintah daerah. Kondisi ini menunjukkan adanya kesenjangan antara ketersediaan data epidemiologis dan pemanfaatannya dalam bentuk model peramalan

yang mampu mengantisipasi perkembangan kasus HIV di masa mendatang.

Pemerintah daerah dan institusi kesehatan di Jawa Barat memerlukan informasi peramalan yang akurat berbasis data historis untuk mendukung perencanaan program pencegahan, alokasi sumber daya kesehatan, dan penguatan kebijakan penanggulangan HIV/AIDS. Namun, penelitian yang secara khusus membandingkan kinerja model peramalan kasus HIV di Jawa Barat menggunakan pendekatan *machine learning* berbasis data terbuka menurut kelompok umur masih terbatas. Oleh karena itu, penelitian ini bertujuan menganalisis secara komparatif kinerja Support Vector Regression (SVR) dan Decision Tree Regression (DTR) dalam meramalkan jumlah kasus HIV di Jawa Barat. Kontribusi penelitian ini adalah penyediaan model peramalan yang tervalidasi secara kuantitatif beserta evaluasi kinerja antar algoritma, yang dapat dimanfaatkan sebagai dasar pengambilan keputusan dan perumusan kebijakan penanggulangan HIV/AIDS berbasis bukti.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Beberapa penelitian untuk meramalkan, atau memprediksi masa depan dengan mengimplementasikan pembelajaran mesin (*machine learning*) khususnya peramalan pada penyakit sudah banyak dilakukan, antara lain:

- Prediksi *Body Mass Index* (BMI) pada dataset Asma dengan komparasi kinerja regresi Decision Tree dan Regresi Linear berganda. Penelitian ini menemukan bahwa Algoritma regresi linear berganda lebih unggul dari decision tree dalam melakukan peramalan [4].
- Prediksi kasus Covid-19 di India menggunakan SVR. Penelitian ini menemukan bahwa algoritma SVR dengan kernel Radial basis function (RBF) menunjukkan performa tinggi, dengan akurasi lebih dari 97% untuk meramalkan kematian, kesembuhan, dan kasus kumulatif, serta 87% untuk kasus harian [5].
- Peramalan jumlah kasus penyakit menular dengan perbandingan SVR, Regresi linier sederhana dan single moving average. Penelitian menemukan metode SVR adalah metode terbaik untuk peramalan HIV dengan akurasi error 36.67% [6].
- Peramalan Prevalensi Stunting di Nusa Tenggara Timur (NTT). Penelitian tersebut menunjukkan bahwa algoritma DTR memiliki kesalahan yang rendah sehingga mampu memberikan estimasi yang baik dalam memprediksi kasus stunting di NTT [7].
- Prediksi penyakit Stroke dengan perbandingan algoritma C4.5 dan SVR. Hasil penelitian menemukan bahwa Decision Tree C4.5 lebih unggul dalam melakukan prediksi daripada model SVR [8].

Berdasarkan berbagai penelitian terdahulu, upaya prediksi di bidang kesehatan telah dilakukan menggunakan beragam algoritma pembelajaran mesin, mulai dari regresi linear berganda, Decision Tree, hingga Support Vector Regression (SVR). Berbagai studi menunjukkan bahwa performa model sangat bergantung pada karakteristik penyakit, struktur dataset, serta kompleksitas pola yang ingin diprediksi. Beberapa penelitian juga menemukan bahwa regresi linear lebih unggul dibandingkan Decision Tree dalam memprediksi BMI pada dataset asma, sementara model berbasis SVR terbukti efektif untuk peramalan kasus COVID-19, dan prediksi penyakit menular tertentu. Di sisi lain, Decision Tree Regression (DTR) menunjukkan performa yang baik dalam memprediksi prevalensi stunting serta penyakit stroke. Temuan-temuan ini menunjukkan bahwa belum ada konsensus mengenai model yang paling sesuai; setiap algoritma memiliki keunggulan pada konteks dan jenis data tertentu. Meskipun penelitian-penelitian tersebut memberikan kontribusi penting dalam pengembangan metode peramalan pada data kesehatan, masih terdapat beberapa celah penelitian (*research gaps*) yang perlu dijembatani. Pertama, sebagian besar studi berfokus pada satu jenis penyakit atau satu jenis model saja, sehingga belum banyak penelitian yang membandingkan performa SVR dan DTR secara langsung dalam konteks prediksi HIV pada tingkat

wilayah. Kedua, penelitian sebelumnya juga belum mengintegrasikan analisis prediksi dengan pendekatan berbasis per kabupaten/kota, padahal pola penyebaran HIV bersifat heterogen antar wilayah. Selain itu, sebagian studi hanya menilai hasil prediksi atau peramalan tanpa menyajikan eksplorasi spasial atau visualisasi komparatif antara data aktual dan data hasil peramalan. Dengan demikian, penelitian ini berupaya mengisi kekosongan tersebut dengan membandingkan kinerja SVR dan DTR secara komprehensif untuk meramalkan jumlah kasus HIV di Jawa Barat sekaligus menghasilkan estimasi pada tingkat wilayah yang lebih detail.

2.2. Landasan Teori

1. Definisi HIV/AIDS

Human Immunodeficiency Virus (HIV) adalah virus yang menyerang sel CD4 pada sistem kekebalan tubuh, sehingga melemahkan kemampuan tubuh melawan infeksi. Penurunan sel CD4 meningkatkan risiko penyakit oportunistik dan, tanpa penanganan, dapat berkembang menjadi AIDS. HIV menular melalui kontak dengan cairan tubuh tertentu, seperti darah, semen, cairan vagina, dan ASI [9].

2. Phyton

Python merupakan bahasa pemrograman yang popular untuk digunakan karena sintaxnya mudah di ingat dan dipahami. Python merupakan open source yang umum digunakan untuk melakukan

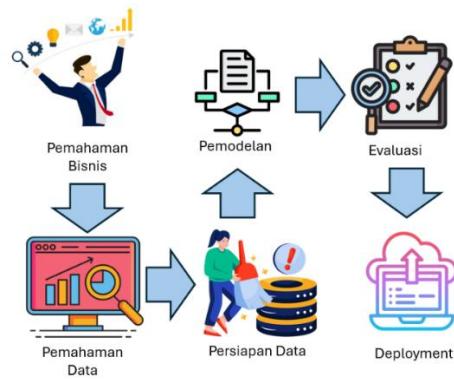
pembuatan website maupun pengolahan data dengan jumlah yang besar, dalam analisis data python dapat melakukan kalkulasi statistik, menganalisa data maupun visualisasi data, Sedangkan pada machine learning python dapat digunakan untuk membuat algoritma untuk modul [10], [11].

3. Data Mining

Data mining adalah salah satu disiplin ilmu yang bertujuan menggali informasi berharga dari dataset yang besar dan kompleks [12].

2.3. Tahapan Penelitian

Alur penelitian menggunakan tahapan yang disajikan pada Gambar 1 dengan menggunakan metode *Cross Industry Standard Process For Data Mining* (CRISP-DM) [13]. Berikut ini adalah penjelasan dari tiap tahapan tersebut.

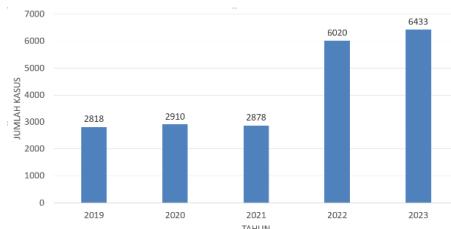


Gambar 1. Tahapan Penelitian Analisis Komperatif untuk Peramalan HIV di Jawa Barat

3. Metode Penelitian

Pada tahap ini, memahami konteks permasalahan kesehatan publik di Jawa Barat terkait

peningkatan jumlah kasus HIV dimana Pada tahun 2022, Jawa Barat berada di perangkat kedua tingkat nasional dengan 1,115 orang yang hidup dengan HIV/AIDS di wilayah tersebut [2].



Gambar 2. Data Jumlah Kasus HIV setelah di bersihkan

3.1. Pemahaman Data

Pada tahap ini, memahami dataset yang bersumber dari open data Jawa Barat yaitu Jumlah Kasus HIV berdasarkan Kelompok Umur di Jawa Barat. Dataset ini berisi kumpulan data jumlah kasus HIV yang terjadi dari berbagai kota dan kabupaten di Jawa Barat dengan total kasus mencapai 31,986 yang terjadi pada tahun 2019 hingga tahun 2023. Pada penelitian ini menggunakan lima atribut yaitu nama_kabupaten_kota, kelompok_umur, jenis_kelamin, jumlah_kasus, tahun.

Tabel 1. Atribut Penelitian

No	Atribut	Type	Kode Atribut
1	nama_kabupaten_kota	nominal	K1
2	kelompok_umur	nominal	K2
3	jenis_kelamin	nominal	K3
4	jumlah_kasus	numerik	K4
5	tahun	numerik	K5

Tabel 1 menampilkan jenis atribut yang digunakan dalam penelitian, diklasifikasikan menjadi nominal

dan numerik, serta mencantumkan kode atribut K1 hingga K5. Sedangkan, pada Tabel 2 menampilkan sampel dari dataset yang digunakan dalam penelitian untuk meramalkan jumlah kasus HIV yang terjadi di Jawa Barat. Laki-Laki bernilai 1, dan Perempuan bernilai 0.

Tabel 2. Sampel dataset yang digunakan

No	K1	K2	K3	K4	K5
1	kabupaten bogor	0-4	1	4	2019
2	kabupaten bogor	0-4	0	7	2019
3	kabupaten bogor	5-14	1	3	2019
...
1617	kota banjar	≥50	0	1	2023

3.2. Persiapan Data

Pada tahap ini, merupakan tahap persiapan data yang dilakukan melalui 2 langkah yaitu menghilangkan data yang duplikasi (*remove duplicate*), menghapus data yang bernilai 0 atau *remove missing value* dan melakukan seleksi data [14]. Semua proses ini menggunakan fitur *sorting* dan *delete* dengan aplikasi Microsoft Excel.

1. Seleksi Data berdasarkan Jumlah Kasus.

Pada tahap ini, melakukan pembersihan data dengan cara menghapus dataset yang memiliki jumlah kasus 0 hingga kurang dari 10. Penghapusan data ini bertujuan menghindari data *overfitting* dan membuat akurasi menjadi lebih optimal.

Tabel 3. Sampel dataset setelah pembersihan

No.	K1	K2	K3	K4	K5
1	kabupaten bogor	20-24	1	48	2019
2	kabupaten bogor	20-24	0	21	2019
3	kabupaten bogor	25-49	1	217	2019
4	kabupaten bogor	25-49	0	133	2019
5	kabupaten bogor	≥50	1	16	2019
...
495	kota banjar	25-49	1	33	2023

Pada Tabel 3 menampilkan sampel dataset setelah proses pembersihan, dan mengalami penyusutan jumlah data.

2. Seleksi Data Kelompok Umur

Langkah selanjutnya seleksi data pada atribut kelompok_umur, dan jumlah_kasus. Pada Penelitian ini, hanya menggunakan kelompok umur yang produktif karena sangat beresiko terhadap penularan HIV/AIDS. Berdasarkan data WHO, Infeksi HIV/AIDS sebagian besar diderita oleh kelompok usia produktif (15-49 tahun), sehingga jumlah kasus pada usia 0-4 dan ≥50 dihapus [15]. Pada Tabel 4, menampilkan sampel dataset hasil seleksi dari kedua tahap.

Tabel 4. Data hasil seleksi kelompok umur

No	K1	K2	K3	K4	K5
1	kabupaten bandung	15-19	1	10	2019
2	kabupaten bandung	20-24	1	30	2019
3	kabupaten bandung	25-49	1	91	2019
4	kabupaten bandung	25-49	0	33	2019
5	kabupaten bekasi	20-24	1	34	2019
...
243	kota depok	25-49	0	63	2023

3.3. Pemodelan Data

Pada tahap ini melakukan pemodelan dengan menggunakan algoritma *machine learning* untuk mengetahui performa terbaik dari algoritma untuk memprediksi jumlah kasus HIV yang terjadi di Jawa Barat. Pada penelitian ini, membandingkan 2 algoritma, yaitu Algoritma *Support Vector Regression* (SVR), dan *Decision Tree Regression* (DTR) menggunakan parameter standar yang sudah dimiliki oleh kedua algoritma tersebut.

Parameter yang digunakan SVR yaitu C, epsilon, kernel. Sedangkan pada algoritma DTR menggunakan `max_depth`, `min_samples_split`, `min_samples_leaf` untuk mengatur keseimbangan antara akurasi peramalan dan kompleksitas algoritma, serta mencegah *overfitting* atau *underfitting* yang berguna mengoptimalkan performa kedua algoritma tersebut.

1. Support Vector Regression (SVR)

Support Vector Regression (SVR) merupakan pengembangan dari metode Support Vector Machines (SVM) yang dirancang untuk menyelesaikan permasalahan regresi. Pendekatan ini bertujuan menemukan suatu fungsi dalam bentuk *hyperplane* yang berperan sebagai fungsi regresi, sehingga mampu merepresentasikan seluruh data masukan dengan tingkat kesalahan tertentu yang diusahakan sekecil mungkin [5]. SVR menggabungkan konsep margin maksimum dan kernel trick seperti pada SVM untuk menangani

regresi. Model ini berupaya menentukan fungsi regresi $f(x)$ yang tetap berada dalam jarak ϵ dari nilai target, sekaligus menjaga fungsi tetap sederhana. Perbedaan peramalan yang masih berada dalam batas ϵ dianggap dapat ditoleransi dan tidak dicatat sebagai *error*. Sebagai contoh terdapat data training, (x_i, y_i) , $i = 1, \dots, l$ dengan data input $x = \{x_1, \dots, x_l\} \subseteq \mathbb{R}^n$ dan $y = \{y_1, \dots, y_l\} \subseteq \mathbb{R}$ dan l merupakan banyaknya data *trining*.

Dengan Fungsi sebagai berikut:

$$(x) = w(x) + b \quad (1)$$

keterangan :

W : vektor pembobot

(x) : fungsi yang memetakan x dalam suatu dimensi

B : bias

2. Decision Tree Regression (DTR)

Decision Tree Regression (DTR) merupakan salah satu metode regresi non-linear yang menggunakan struktur pohon keputusan untuk memprediksi nilai output berdasarkan nilai input [16]. Metode ini bekerja dengan cara membagi (*splitting*) data ke dalam beberapa node berdasarkan kondisi tertentu agar menghasilkan kelompok data (*leaf*) yang memiliki kemiripan nilai output paling tinggi. DTR tidak memerlukan asumsi linearitas maupun skala data yang besar karena pendekatan utamanya bersifat berbasis aturan (*rule-based*). Setiap percabangan (*branch*) dalam pohon merepresentasikan keputusan berdasarkan fitur tertentu, sedangkan *leaf node*

menunjukkan nilai prediksi akhir.

Dengan demikian, model ini berfungsi untuk meminimalkan kesalahan peramalan dengan cara memilih pembagian (*split*) yang menghasilkan penurunan impurity paling besar. Sebagai contoh, terdapat data training (x_i, y_i) , $i = 1, \dots, n$, dengan data input $x = \{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$ dan $y = \{y_1, \dots, y_n\} \subseteq \mathbb{R}$. Model akan membagi ruang fitur menjadi beberapa region R_1, R_2, \dots, R_m , lalu menentukan nilai peramalan rata-rata pada setiap region tersebut. Dengan fungsi sebagai berikut:

$$f(x) = \sum_{j=1}^m c_j I(x \in R_j) \quad (2)$$

Keterangan:

R_j : region atau area hasil pemisahan (*split*) data

c_j : nilai peramalan (biasanya rata-rata dari nilai y_i di dalam R_j)

$I(x \in R_j)$: fungsi indikator, bernilai 1 jika x berada di region R_j , dan 0 jika tidak

m : jumlah leaf node atau region akhir.

3.4. Evaluasi

Pada tahap ini, melakukan evaluasi performa terhadap akurasi dan efektifitas hasil dari 2 algoritma regresi yaitu algoritma SVR, dan DTR. Evaluasi ini menggunakan beberapa model evaluasi menurut Chicco, diantaranya *Coefficient of Determination* (R^2) yang digunakan untuk mengukur seberapa besar variasi data aktual dapat dijelaskan oleh model [17], [18], *Root Mean*

Squared Error (RMSE) merupakan akar dari rata-rata kuadrat selisih antara nilai aktual dan prediksi, menunjukkan seberapa besar rata-rata kesalahan peramalan dalam satuan yang sama dengan data asli.

Selain itu menggunakan *Mean Absolute Percentage Error* (MAPE) merupakan rata-rata persentase kesalahan absolut antara nilai aktual dan prediksi, memberi gambaran seberapa besar *error* dalam persen. Penggunaan matriks R^2 menilai kecocokan model, RMSE menilai akurasi absolut, dan MAPE menilai akurasi relatif (persentase kesalahan) sehingga memberikan informasi seberapa jauh model yang sedang digunakan menyimpang dari nilai aktual dan seberapa akurat model yang digunakan^[17].

3.5. Deployment

Tahap ini menyusun hasil eksperimen dengan membandingkan performa SVR dan DTR serta menampilkan peramalan kasus HIV di Jawa Barat untuk tahun berikutnya per Kabupaten atau Kota.

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

Evaluasi kinerja dilakukan untuk membandingkan efektivitas algoritma Support Vector Regression (SVR) dan Decision Tree Regression (DTR) dalam meramalkan jumlah kasus HIV di Jawa Barat.

Proses pembersihan dan normalisasi data

dilakukan untuk menjaga konsistensi variabel, kemudian dataset dibagi menjadi data latih (80%) dan data uji (20%) guna memastikan kemampuan generalisasi model.

Pendekatan ini memungkinkan pengukuran akurasi peramalan secara objektif pada periode data yang tidak digunakan saat pelatihan model. Hasil pengujian yang ditampilkan pada Tabel 5 menunjukkan bahwa DTR memiliki performa lebih baik dibandingkan SVR berdasarkan tiga metrik utama, yaitu R^2 , RMSE, dan MAPE. Algoritma DTR menghasilkan R^2 sebesar 83.62, RMSE sebesar 32.29, dan MAPE sebesar 33.3, sedangkan SVR memberikan hasil yang lebih rendah.

Temuan tersebut diperkuat oleh nilai rata-rata performa pada Tabel 6, yang secara konsisten menempatkan DTR sebagai model dengan kinerja paling optimal. Dengan demikian, dapat disimpulkan bahwa Decision Tree Regression lebih efektif digunakan untuk meramalkan jumlah kasus HIV di Jawa Barat.

Tabel 5. Perbandingan Nilai Rata-Rata Metrik Kinerja antara SVR dan DTR

	SVR	DTR
R^2	36.33	83.62
RMSE	84.01	32.29
MAPE	44.87	33.30

a. Coefficient of Determination (R^2)

Berdasarkan Tabel 5, model SVR hanya mampu menjelaskan 36.33% variasi data, sehingga kurang efektif dalam menangkap hubungan

antara variabel input dan target. Sebaliknya, model DTR menunjukkan kinerja yang lebih unggul dengan nilai R^2 sebesar 83.62%, yang mengindikasikan kemampuan yang lebih baik dalam merepresentasikan pola data serta kesesuaian yang lebih tinggi terhadap karakteristik dataset.

b. Root Mean Squared Error (RMSE)

Berdasarkan Tabel 5, Model SVR menghasilkan nilai RMSE sebesar 84.01, yang menunjukkan kesalahan peramalan rata-rata mencapai 84 satuan terhadap nilai aktual. Nilai ini tergolong besar dan menunjukkan kemampuan generalisasi yang rendah. Sementara itu, model DTR unggul dengan memperoleh nilai RMSE sebesar 32.29, yang berarti kesalahan peramalan rata-rata hanya sekitar 32 satuan. Nilai yang jauh lebih kecil ini mengindikasikan bahwa DTR mampu meramalkan dengan lebih akurat dan stabil.

c. Mean Absolute Percentage Error (MAPE)

Berdasarkan Tabel 5, Model SVR mencapai nilai MAPE sebesar 44.87, yang menunjukkan bahwa prediksi model rata-rata menyimpang hampir setengah dari nilai aktual. Sebaliknya, model DTR lebih unggul dengan menghasilkan nilai MAPE sebesar 33.30, dengan tingkat penyimpangan yang lebih rendah. Meskipun selisih antara kedua model tidak terlalu besar, terdapat peningkatan efisiensi.

Tabel 6. Hasil Perbandingan Kinerja Algoritma SVR dan DTR

Kabupaten/ Kota	R^2		RMSE		MAPE	
	SV R	DT R	SVR R	DT R	SV R	DT R
kota bogor	79. 45	94. 49	29.5 7	15. 32	39. 16	37. 27
kabupaten indramayu	38. 93	94. 31	53.5 6	16. 34	74. 54	50. 56
kabupaten bogor	46. 99	93. 14	106. 02	38. 14	34. 2	17. 78
kota bandung	24. 15	92. 58	223. 53	69. 91	63. 67	47. 00
kota bekasi	31. 19	86. 61	48.2 8	21. 29	33. 38	22. 99
kota cirebon	-	40. 3	43.0 8	32. 76	24. 27	24. 18

Nilai yang ditebalkan adalah yang terbaik

3.6. Deployment

Berdasarkan hasil evaluasi pada Tabel 5 yang menunjukkan bahwa algoritma DTR memiliki kinerja lebih unggul dibandingkan SVR, maka tahap *deployment* berfokus pada penyajian perbandingan antara data aktual dan hasil peramalan, serta analisis keduanya menggunakan model DTR.

a. Perbandingan Data Aktual Dan Estimasi

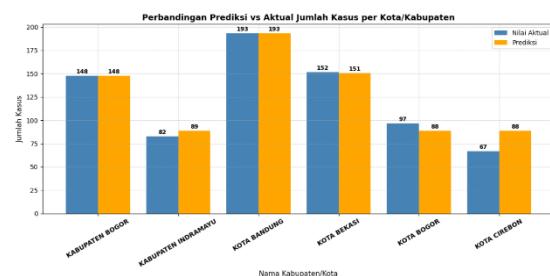
Tabel 7 menyajikan perbandingan antara data aktual dan estimasi per tahun yang dihasilkan oleh algoritma DTR.

Tabel 7. Perbandingan Data Aktual dan Data Estimasi Per-Tahun

Tahun	Data Aktual	Data Estimasi
2019	2818	2777
2020	2910	2946
2021	2878	2877
2022	6020	6150
2023	6433	6309
Jumlah	21059	21059
Rata-Rata	4211.8	4211.8

b. Peramalan Jumlah Kasus HIV Tahun 2024

Gambar 3 menyajikan visualisasi hasil peramalan jumlah kasus HIV di Jawa Barat untuk tahun 2024 dengan menggunakan algoritma DTR sebagai model dengan performa paling optimal. Estimasi tersebut diperoleh berdasarkan data kasus tahun 2023 yang telah melalui proses seleksi pada tahap persiapan dataset.



Gambar 3. Visualisasi Perbandingan Data Aktual dan Prediksi atau Estimasi pada tahun 2024

Dari data grafik batang peramalan tahun 2024 terhadap jumlah kasus HIV di provinsi Jawa Barat menunjukkan bahwa kota Bandung diramalkan memiliki jumlah kasus yang paling tinggi dibandingkan dengan kota/kabupaten lainnya. Hasil peramalan ini dapat menjadi acuan pemerintah atau *stakeholder* untuk mengambil kebijakan yang tepat dan sedini mungkin untuk mengatasi penyebaran penyakit HIV di provinsi Jawa Barat dengan memfokuskan pada kota Bandung, kemudian dilanjutkan pada kota Bekasi yang menduduki ke-2 tertinggi, kemudian kabupaten Bogor dan kota Bogor yang tidak kalah tinggi serta kabupaten Indramayu dan kota Cirebon.

5. Kesimpulan

Penelitian ini membandingkan kinerja algoritma Support Vector Regression (SVR) dan Decision Tree Regression (DTR) dalam meramalkan jumlah kasus HIV di Provinsi Jawa Barat. Berdasarkan tiga metrik evaluasi yaitu R^2 , RMSE, dan MAPE, maka diketahui bahwa model DTR menunjukkan performa yang lebih unggul dengan nilai R^2 sebesar 83.62, RMSE sebesar 32.29, dan MAPE sebesar 33.30. Sementara itu, SVR menghasilkan performa yang lebih rendah dengan nilai R^2 sebesar 36.33, RMSE sebesar 84.01, dan MAPE sebesar 44.87. Dengan demikian, Decision Tree Regression disimpulkan sebagai model yang lebih efektif dan akurat untuk meramalkan jumlah kasus HIV di Jawa Barat. Penelitian di masa mendatang, disarankan untuk menguji algoritma ensemble, seperti Random Forest atau XGBoost, guna mengeksplorasi potensi peningkatan akurasi dan stabilitas model prediksi.

6. Daftar Pustaka

- [1] Kemenkes, "Berani Tes, Berani Lindungi Diri, Kemenkes Targetkan Eliminasi HIV dan IMS Tahun 2030," <https://kemkes.go.id/>, 2025.
- [2] N. S. Surjandari *et al.*, "Sosialisasi Pencegahan HIV/AIDS dan Penyalahgunaan Narkoba di Desa Winong, Kecamatan Mirit, Kabupaten Kebumen," *Kreasi J. Inov. dan Pengabd. Kpd. Masy.*, vol. 2, no. 2, pp. 219–225, 2022.
- [3] P. Taniwan, A. F. Bilbina, C. R. S. Ganap, and D. Y. Faidah, "Pemodelan Faktor-

- Faktor yang memengaruhi Kejadian HIV/AIDS di Provinsi Jawa Barat," *IMEIJ Indo-mathEdu Intellectuals J.*, vol. 5, no. 3, pp. 3298–3308, 2024.
- [4] A. T. Nurani, A. Setiawan, and B. Susanto, "Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma," *J. Sains dan Edukasi Sains*, vol. 6, no. 1, pp. 34–43, 2023.
- [5] D. Parbat and M. Chakraborty, "A python based support vector regression model for prediction of COVID19 cases in India," *Chaos, Solitons & Fractals*, vol. 138, no. 109942, pp. 1–5, 2020, doi: 10.1016/j.chaos.2020.109942.
- [6] M. T. Siregar, G. M. A. Sasmita, and G. A. A. Putri, "Perbandingan Analisis Metode Peramalan Jumlah Kasus Penyakit Menular di Kota Bandung (Studi Kasus: Dinas Kesehatan Kota Bandung)," *JITTER J. Ilm. Teknol. dan Komput.*, vol. 3, no. 1, pp. 831–842, 2022.
- [7] I. S. I. Putri, R. S. Pradini, and M. Anshori, "Decision Tree Regression untuk Prediksi Prevalensi Stunting di Provinsi Nusa Tenggara Timur," *J. Teknologi Inform. dan Komput. MH. Thamrin*, vol. 10, no. 2, pp. 413–427, 2024.
- [8] F. Akbar, H. W. Saputra, A. K. Maulaya, M. F. Hidayat, and R. Rahmaddeni, "Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 61–67, 2022.
- [9] H. A. F. Sinaga, "HIV/AIDS," *Med. Lab. J.*, vol. 3, no. 2, pp. 45–52, 2025.
- [10] B. A. C. Permana, M. Sadali, and R. Ahmad, "Penerapan Model Decision Tree Menggunakan Python Untuk Prediksi Faktor Dominan Penyebab Penyakit Stroke," *J. Inform. dan Teknol.*, vol. 7, no. 1, pp. 23–31, 2024, [Online]. Available: <https://e-journal.hamzanwadi.ac.id/index.php/infotek/article/view/23232>.
- [11] B. A. C. Permana and M. Djamaruddin, "Penerapan Python Dalam Data Mining Untuk Prediksi Kangker Paru," *Infotek J. Inform. dan Teknol.*, vol. 6, no. 2, pp. 470–477, 2023, doi: 10.29408/jit.v6i2.17816.
- [12] R. F. Putra et al., *DATA MINING: Algoritma dan Penerapannya*. Jambi: PT. Sonpedia Publishing Indonesia, 2023.
- [13] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi:<https://doi.org/10.1016/j.procs.2021.01.199>.
- [14] I. Koumarelas, L. Jiang, and F. Naumann, "Data Preparation for Duplicate Detection," *J. Data Inf. Qual.*, vol. 12, no. 3, pp. 1–24, 2020, doi: 10.1145/3377878.
- [15] WHO, "Prevalence of HIV among adults aged 15 to 49 (%)," <https://www.who.int/>, 2025. [\(accessed Jan. 01, 2024\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-hiv-among-adults-aged-15-to-49-(-)).
- [16] M. Jena and S. Dehuri, "Decision Tree for Classification and Regression : A State-of-the Art Review," *Informatica*, vol. 44, pp. 405–420, 2020.
- [17] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient ofdetermination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, no. e623, pp. 1–24, 2021, doi: 10.7717/peerj-cs.623.
- [18] H. Hewamalage, K. Ackermann, and C. Bergmeir, "Forecast evaluation for data scientists: common pitfalls and best practices," *Data Min. Knowl. Discov.*, vol. 37, no. 2, pp. 788–832, 2023, doi: 10.1007/s10618-022-00894-