Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

Analisis Kinerja Model Support Vector Machine dalam Prediksi Kasus HIV di Indonesia Berdasarkan Data Time Series

Muhammad Al-Ghifari Erza, Putri Taqwa Prasetyaningrum*

Fakultas Teknologi Informasi, Sistem informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia
Email: [1]221220081@student-mercubuana-yogya.ac.id, [2,*]putri@mercubuana-yogya.ac.id
Email Penulis Korespondensi:putri@mercubuana-yogya.ac.id
Submitted: **16/05/2025**; Accepted: **16/06/2025**; Published**: 20/06/2025**

Abstrak−Prediksi jumlah kasus HIV yang akurat sangat penting dalam upaya pengendalian epidemi yang efektif di Indonesia. Seiring meningkatnya jumlah kasus dan kompleksitas faktor penyebaran, metode prediksi berbasis machine learning semakin relevan digunakan. Penelitian ini menganalisis performa model Support Vector Machine (SVM) dalam meramalkan jumlah kasus HIV di Indonesia menggunakan data time series dari tahun 2012 hingga 2024. Metode CRISP-DM digunakan sebagai kerangka kerja dalam proses analisis, dimulai dari pemahaman bisnis hingga deployment model. Dataset yang digunakan mencakup data sekunder dari Kementerian Kesehatan, seperti SIHA, surveilans nasional, dan laporan Ditjen P2P. Model SVM dipilih karena kemampuannya dalam menangani data non-linier dan jumlah data yang terbatas, serta ketahanannya terhadap overfitting. Evaluasi model dilakukan menggunakan metrik MAE, RMSE, dan MAPE. Hasil penelitian menunjukkan bahwa model SVR dengan kernel RBF memberikan akurasi prediksi yang baik, dengan nilai MAE sebesar 691,34, RMSE sebesar 823,11, dan MAPE sebesar 13% pada data pengujian. Dengan demikian, SVM dapat menjadi alat yang efektif untuk mendukung pengambilan keputusan berbasis data dalam pengendalian HIV di Indonesia.

Kata Kunci: Support Vector Machine; Prediksi Epidemi; HIV; Time Series; Machine Learning

Abstract−Accurate predictions of HIV cases are crucial in efforts to control the epidemic effectively in Indonesia. As the number of cases and the complexity of transmission factors increase, machine learning-based prediction methods are becoming increasingly relevant. This study analyzes the performance of the Support Vector Machine (SVM) model in forecasting the number of HIV cases in Indonesia using time series data from 2012 to 2024. The CRISP-DM methodology is used as the framework for the analysis process, starting from business understanding to model deployment. The dataset used includes secondary data from the Ministry of Health, such as SIHA, national surveillance, and reports from the Directorate General of Disease Prevention and Control (Ditjen P2P). The SVM model is selected due to its ability to handle non-linear data and limited data sizes, as well as its resilience to overfitting. Model evaluation is performed using MAE, RMSE, and MAPE metrics. The results of the study show that the SVR model with an RBF kernel provides good prediction accuracy, with MAE values of 691.34, RMSE of 823.11, and MAPE of 13% on the test data. Therefore, SVM can be an effective tool to support data-driven decision-making in HIV control efforts in Indonesia.

Keywords Support Vector Machine; Epidemic Prediction; HIV/AIDS; Time Series; Machine Learning

1. PENDAHULUAN

Human Immunodeficiency Virus (HIV) dan *Acquired Immunodeficiency Syndrome* (AIDS) tetap menjadi tantangan serius bagi kesehatan masyarakat di Indonesia. Data dari Kementerian Kesehatan Republik Indonesia menunjukkan bahwa jumlah kumulatif kasus HIV yang dilaporkan hingga Desember 2023 mencapai 566.707 kasus, dengan peningkatan signifikan terjadi pada tahun 2023 sebanyak 57.299 kasus baru. Penyebaran HIV/AIDS tidak hanya terbatas pada kelompok risiko tinggi seperti pekerja seks dan pengguna narkoba suntik, tetapi juga meluas ke populasi umum, terutama di wilayah perkotaan dengan mobilitas tinggi[1].

Prediksi jumlah kasus HIV/AIDS yang akurat sangat penting dalam mendukung kebijakan kesehatan publik yang proaktif[2]. Dengan kemampuan memproyeksikan tren penyebaran, prediksi yang akurat dapat membantu perencanaan distribusi sumber daya, penentuan prioritas program intervensi, serta evaluasi efektivitas kebijakan sebelumnya. Oleh karena itu, pemodelan prediktif yang mampu menangkap dinamika kompleks dari penyebaran HIV/AIDS menjadi kebutuhan yang mendesak dalam konteks pengambilan keputusan berbasis bukti [3].

Selama ini, metode konvensional seperti regresi linier, *moving average,* dan *Autoregressive Integrated Moving Average* (ARIMA) telah banyak digunakan untuk peramalan kasus penyakit menular[4]. Namun, pendekatan tersebut memiliki keterbatasan dalam menangani karakteristik data epidemiologi yang bersifat *nonlinier*, dinamis, dan dipengaruhi oleh berbagai faktor eksternal yang saling berkaitan [5].

Sebagaimana yang dijelaskan dalam *Foundations of Epidemiology*, epidemiologi adalah ilmu yang berfokus pada distribusi dan determinan penyakit dalam populasi[6]. Dalam konteks HIV/AIDS, pemahaman mendalam mengenai faktor-faktor yang memengaruhi penyebaran penyakit ini sangat penting, terutama mengingat dinamika kompleks yang memengaruhi pola penularan di masyarakat. Buku ini juga menekankan pentingnya penggunaan metode prediktif dalam merencanakan dan mengevaluasi program intervensi kesehatan yang berbasis bukti, yang sejalan dengan penerapan algoritma *machine learning* seperti *Support Vector Machine* (SVM) dalam memprediksi tren penyebaran HIV/AIDS.

Dalam beberapa tahun terakhir, pendekatan machine learning mulai banyak diterapkan dalam bidang epidemiologi karena kemampuannya dalam memproses data besar dan kompleks [7]. Salah satu algoritma yang menonjol adalah *Support Vector Machine* (SVM), yang dikenal efektif dalam menyelesaikan masalah regresi dan klasifikasi, khususnya untuk data berdimensi tinggi dan tidak linier[8]. Dengan menggunakan teknik kernel, SVM

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

mampu memproyeksikan data ke ruang berdimensi lebih tinggi untuk menangkap pola yang tersembunyi dan tidak linear[9] Selain itu, SVM memiliki keunggulan dalam mengurangi risiko overfitting dan bekerja dengan baik meskipun data yang tersedia terbatas atau mengandung noise [10], [11].

Beberapa studi internasional telah membuktikan efektivitas SVM dalam meramalkan perkembangan penyakit menular, termasuk HIV/AIDS, dengan tingkat akurasi yang tinggi dibandingkan model tradisional seperti ARIMA dan regresi. Misalnya, penelitian oleh Nisa et al juga menerapkan SVM untuk memantau indikator klinis HIV/AIDS di Asia dan memperoleh akurasi yang tinggi. Namun, penelitian tersebut lebih fokus pada data klinis pasien dan belum mengeksplorasi potensi pemodelan berdasarkan layanan sistem kesehatan masyarakat secara agregat (makro-level), seperti layanan konseling, jumlah tes, dan cakupan ARV [12], [13]. Namun, hingga kini penerapan SVM dalam konteks epidemiologi HIV/AIDS di Indonesia masih sangat terbatas. Minimnya studi lokal yang mengadopsi pendekatan ini menciptakan celah dalam literatur, sekaligus peluang untuk memberikan kontribusi ilmiah dan praktis yang signifikan [14].

Penelitian ini bertujuan untuk menganalisis performa model SVM dalam meramalkan jumlah kasus HIV/AIDS di Indonesia berdasarkan data time series. Dengan menggunakan data sekunder dari sistem pelaporan nasional, model dikembangkan dan dievaluasi menggunakan metrik statistik seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). Diharapkan, hasil dari penelitian ini dapat menjadi dasar bagi pengambil kebijakan dalam merancang strategi pencegahan dan pengendalian HIV/AIDS yang lebih akurat, adaptif, dan berbasis teknologi [15].

Untuk meningkatkan akurasi dalam pemodelan epidemiologi HIV/AIDS, berbagai studi terkini secara konsisten menunjukkan keunggulan pendekatan machine learning, khususnya SVM, jika dibandingkan dengan metode tradisional. mengungkapkan bahwa SVM sangat efisien dalam mengolah data time series terkait penyakit menular, dengan kemampuan untuk mendeteksi pola yang tersembunyi dalam data yang kompleks dan non-linier[16]. dengan menggunakan teknik feature selection berbasis wrapper pada data HIV di Ethiopia, dan menemukan bahwa penggabungan teknik seleksi fitur dengan SVM dapat meningkatkan kinerja prediksi secara signifikan[17]. Penelitian oleh Abade et al. (2024) dalam studi komparatif mereka, menunjukkan bahwa model berbasis machine learning seperti Support Vector Machine (SVM) memberikan performa lebih baik dibandingkan metode klasik seperti ARIMA dalam kasus prediksi ko-infeksi HIV/TB. Mereka menemukan bahwa SVM lebih tahan terhadap noise dan mampu menghasilkan prediksi yang stabil pada data time series yang terbatas. Meskipun demikian, konteks penelitian mereka masih terbatas pada kasus internasional dan belum banyak menyentuh aplikasi di negara berkembang seperti Indonesia.[18].

Penelitian lain juga menggaris bawahi pentingnya eksplorasi model SVM di konteks negara berkembang. menerapkan data mining dalam sistem surveilans HIV/AIDS dan menunjukkan bahwa SVM dapat memberikan sistem pendukung keputusan yang efisien untuk perencanaan intervensi kesehatan[19]. Oleh karena itu meskipun model deep learning seperti LSTM populer, SVM tetap relevan untuk data jangka pendek dan prediksi insidensi AIDS bulanan dengan sumber daya terbatas[20].

Dengan demikian, terdapat beberapa kesenjangan penting (research gap) dalam literatur yang menjadi latar belakang utama penelitian ini. Pertama, masih minimnya studi lokal yang secara khusus mengadopsi pendekatan machine learning, terutama Support Vector Machine (SVM), untuk memprediksi kasus HIV di Indonesia. Sebagian besar penelitian terdahulu masih menggunakan metode konvensional seperti regresi linier atau ARIMA, yang kurang mampu menangani pola non-linier dan kompleksitas temporal dalam data epidemiologi. Kedua, belum banyak penelitian yang secara eksplisit mengeksplorasi pemanfaatan data layanan kesehatan seperti jumlah tes HIV, layanan konseling, dan pengobatan ARV sebagai fitur prediktor utama dalam model prediksi berbasis time series. Padahal, data layanan ini memiliki potensi tinggi untuk mencerminkan dinamika penyebaran HIV di lapangan. Ketiga, belum terdapat kajian yang mengintegrasikan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) secara menyeluruh dalam pembangunan model prediktif epidemiologi di Indonesia, terutama dalam konteks implementasi kebijakan kesehatan berbasis data. Oleh karena itu, penelitian ini hadir untuk mengisi kekosongan tersebut dengan membangun model prediksi kasus HIV menggunakan SVM, memanfaatkan data layanan kesehatan sebagai variabel input, dan menerapkan pendekatan CRISP-DM secara sistematis dari tahap pemahaman bisnis hingga *deployment*.

2. METODOLOGI PENELITIAN

Penelitian ini mengadopsi pendekatan eksploratif-prediktif dengan memanfaatkan metode machine learning, khususnya *Support Vector Machine* (SVM), untuk meramalkan jumlah kasus HIV di Indonesia. Model prediksi dibangun menggunakan data time series epidemiologi HIV yang dikumpulkan dari berbagai sumber resmi nasional. Proses penelitian mengikuti kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*), yang terdiri dari enam tahap utama: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi model, dan penerapan model [21].

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

2.1. *Business Understanding* (Pemahaman Masalah)

Indonesia menghadapi tantangan besar dalam mengendalikan penyebaran HIV. Meningkatnya jumlah kasus dari tahun ke tahun menunjukkan bahwa pendekatan tradisional berbasis pelaporan dan pengamatan saja tidak memadai. Oleh karena itu, tujuan utama dari studi ini adalah membangun model prediksi berbasis machine learning yang mampu memperkirakan jumlah kasus HIV secara akurat setiap tahunnya. Model ini diharapkan dapat digunakan untuk mendukung proses perencanaan strategis dan pengambilan keputusan oleh instansi pemerintah, khususnya dalam penyusunan kebijakan kesehatan berbasis data.

2.2. *Data Understanding* (Pemahaman Data)

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari beberapa sumber resmi, yaitu:
a. SIHA (Sistem Informasi HIV-AIDS dan IMS)
b. Laporan dari Direktorat Jenderal Pencegahan dan Pengendalian Penyakit (Ditjen P2P)
c. Data Surveilans HIV Nasional

Dataset mencakup periode tahun 2012 hingga 2024 dengan cakupan data triwulanan. Fitur-fitur yang dikumpulkan antara lain jumlah kasus HIV baru, jumlah tes HIV yang dilakukan, jumlah pasien yang menerima terapi antiretroviral (ARV), cakupan layanan konseling, serta indikator dukungan layanan kesehatan lainnya. Hasil eksplorasi awal menunjukkan adanya tren musiman, pola linear, dan kemungkinan efek lag temporal yang mempengaruhi jumlah kasus

2.3. *Data Preparation* (Persiapan Data)

Persiapan data dilakukan dalam beberapa tahap penting:
a. Pembersihan Data (Cleaning):
   Jika terdapat missing values ketika data di import ke workspace, maka akan ditangani menggunakan metode interpolasi linear atau forward fill
b. Rekayasa Fitur (Feature Engineering)
   1. Membuat Time Based Feature
   2. Membuat fitur lag
   3. Menentukan variabel prediktor [layanan_konseling_tes, tes, layanan_pengobatan, pasien_arv]
c. Transformasi Time Series:
   Dataset diubah ke dalam bentuk supervised learning dengan lag features, misalnya:
   1. Input (X): berisi fitur seperti [layanan_konseling_tes, tes, layanan_pengobatan, pasien_arv] dan fitur lag [lag_1 sampai lag_12]
   2. Target (y): variable target yaitu [hiv_nasional]
d. Normalisasi:
   Seluruh fitur dinormalisasi menggunakan MinMaxScaler untuk menjaga kestabilan model karena SVM sangat sensitif terhadap skala fitur.
e. Pembagian Data:
   Data dibagi menjadi:
   1. Training set: 80%
   2. Testing set: 20%
   Validasi dilakukan menggunakan metode Time Series Split , agar memperhitungkan urutan temporal.

2.4. *Modeling* (Pemodelan)

Proses pemodelan dilakukan dengan menggunakan Support Vector Regression (SVR) dari pustaka Scikit-learn di Python. SVR dipilih karena kemampuannya dalam menangani data nonlinier dan jumlah data yang terbatas. Model ini memetakan input ke ruang berdimensi tinggi dan mencari hyperplane terbaik yang dapat meramalkan nilai keluaran dalam margin kesalahan tertentu (epsilon-insensitive loss).

2.5. *Evaluation* (Evaluasi Model)

Evaluasi model dilakukan dengan membandingkan hasil prediksi dengan nilai aktual menggunakan tiga metrik utama:
a. Mean Absolute Error (MAE): Mengukur kesalahan absolut rata-rata.
b. Root Mean Squared Error (RMSE): Menekankan kesalahan besar secara kuadrat.
c. Mean Absolute Percentage Error (MAPE): Memberikan perspektif persentase kesalahan relatif.

Nilai error yang rendah pada ketiga metrik menjadi indikator keberhasilan model dalam memprediksi tren penyebaran HIV/AIDS

2.6. *Deployment* (Penerapan Model)

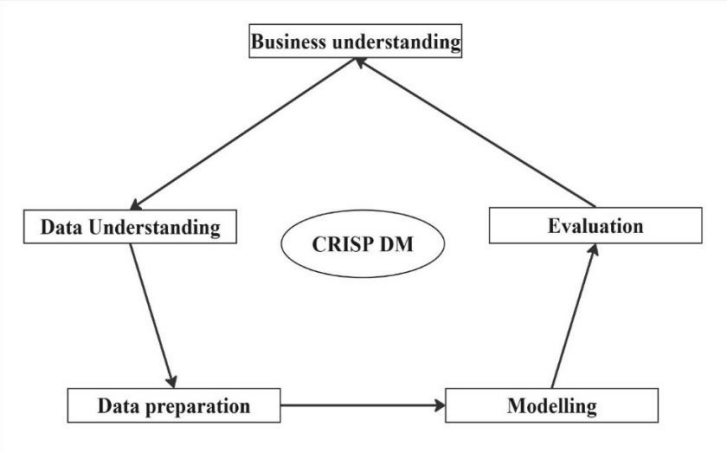
Sebagai tahap akhir, prototipe sistem prediksi dikembangkan menggunakan Python, dilengkapi dengan dokumentasi dan visualisasi hasil prediksi. Meskipun belum diimplementasikan ke dalam sistem nasional, prototipe ini membuka peluang untuk integrasi dengan dashboard monitoring HIV/AIDS berbasis web. Rencana pengembangan lanjutan

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

melibatkan penambahan fitur eksternal seperti faktor demografi, perilaku risiko, dan mobilitas populasi untuk meningkatkan akurasi prediksi.

2.7. Diagram alir

Diagram alir dalam konteks penelitian ini yang menggunakan metode CRISP-DM dan pendekatan prediksi menggunakan model Support Vector Machine (SVM) dapat dilihat pada Gambar 1.



Gambar 1. Diagram alir

Pada Gambar 1 menggambarkan alur proses penelitian yang digunakan dalam studi ini. Proses dimulai dengan pemahaman terhadap tujuan bisnis yang diinginkan, dilanjutkan dengan pengumpulan data, pengolahan data, pemodelan hubungan antar data, evaluasi, dan akhirnya pembuatan laporan yang merangkum hasil temuan dari proses analisis data mining. Alur ini menggunakan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang terdiri dari enam tahap: *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment.*

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dan pembahasan dari penelitian yang telah dilakukan, termasuk penerapan metode *Support Vector Machine* (SVM) dalam memprediksi jumlah kasus HIV di Indonesia. Analisis dilakukan berdasarkan data time series yang diperoleh dari sumber resmi nasional, dengan tujuan untuk mengevaluasi kinerja model dalam konteks epidemiologi HIV/AIDS di Indonesia. Hasil dan pembahasan meliputi 6 tahap sesuai dengan rangkaian kerja CRISP-DM (Gambar 1) dimana pada tahap ini akan mengekstraksi semua informasi yang didapat dari dataset untuk memaksimalkan kinerja model.

3.1 Hasil

3.2.1 Data Understanding

Setelah dilakukan penginputan dataset ‘data hiv perbulan’ ke workspace Googe Colab, dapat diketahui bahwa tidak ada missing values pada data. Pada Gambar 2, menampilkan hasil proses impor dataset ke lingkungan kerja (workspace) Google Colab. Terlihat data terdiri dari 156 baris dan 11 kolom, dengan kolom ‘tahun’ sebagai indeks waktu dan beberapa variabel lain seperti ‘hiv_nasional’, ‘layanan_konseling_tes’, ‘tes’, ‘layanan_pengobatan’, dan ‘pasien_arv’. Gambar ini juga menegaskan bahwa data sudah bersih dari missing values, sehingga siap digunakan untuk analisis lanjutan.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 156 entries, 0 to 155
Data columns (total 11 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   tahun                  156 non-null    datetime64[ns]
 1   hiv_nasional           156 non-null    int64
 2   hiv_sumatera           156 non-null    int64
 3   hiv_jawa               156 non-null    int64
 4   hiv_kalimantan         156 non-null    int64
 5   hiv_sulawesi           156 non-null    int64
 6   hiv_maluku_papua       156 non-null    int64
 7   layanan_konseling_tes  156 non-null    int64
 8   tes                    156 non-null    int64
 9   layanan_pengobatan     156 non-null    int64
 10  pasien_arv             156 non-null    int64
dtypes: datetime64[ns](1), int64(10)
memory usage: 13.5 KB
None
       tahun  hiv_nasional  hiv_sumatera  ...    tes  layanan_pengobatan  pasien_arv
0 2012-01-01          1857           231  ...  21415                  65        7215
1 2012-02-01          2115           301  ...  30415                  87        9315
2 2012-03-01          2019           301  ...  30415                  85        9287
3 2012-04-01          1155           171  ...  14215                  93        7715
4 2012-05-01          1477           221  ...  21015                 117        9735

[5 rows x 11 columns]
```

Gambar 2. Hasil Import data dan Deskripsi data

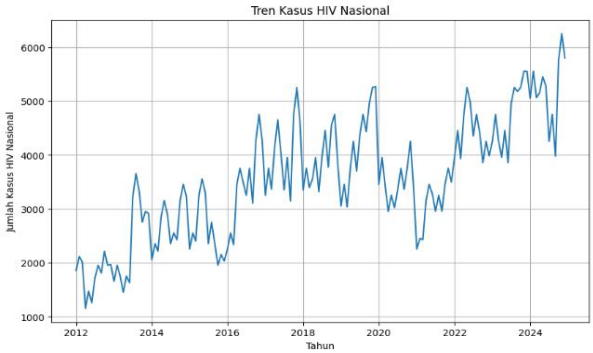
Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

Berdasarkan hasil analisis dan penerapan model Support Vector Machine (SVM) (Gambar 2) dalam memprediksi jumlah kasus HIV/AIDS di Indonesia, dapat disimpulkan bahwa model SVM memberikan performa yang baik dan efektif dalam menghadapi data yang bersifat nonlinier dan dinamis. Penggunaan data time series dari periode 2012 hingga 2024 memungkinkan model ini untuk menangkap pola tren penyebaran HIV/AIDS dengan tingkat akurasi yang lebih tinggi dibandingkan metode tradisional seperti regresi linier dan ARIMA.

Variabel-variabel layanan kesehatan seperti layanan konseling, tes HIV, dan pengobatan antiretroviral terbukti berpengaruh signifikan terhadap prediksi jumlah kasus, yang menunjukkan hubungan kompleks antar faktor epidemiologi yang berhasil dimodelkan oleh SVM. Model ini juga mampu menangani keterbatasan jumlah data serta mengurangi risiko overfitting, sehingga menghasilkan prediksi yang stabil dan andal.

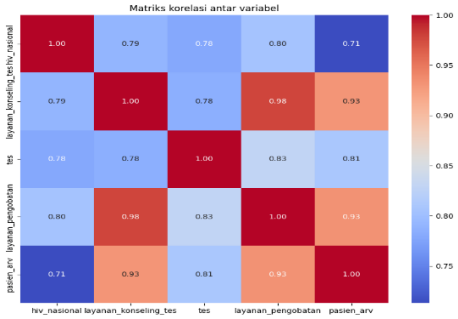
Hasil evaluasi menggunakan metrik MAE, RMSE, dan MAPE mengindikasikan bahwa model SVR dengan kernel RBF memiliki tingkat kesalahan yang rendah dan dapat diandalkan, baik pada data pelatihan maupun pengujian. Dengan demikian, model ini berpotensi menjadi alat bantu yang efektif bagi pengambil kebijakan dalam merancang strategi pencegahan dan pengendalian HIV/AIDS yang berbasis data.

Penelitian ini membuka peluang untuk pengembangan lebih lanjut, seperti integrasi data eksternal yang lebih beragam, peningkatan fitur, serta implementasi sistem prediksi yang terintegrasi dalam monitoring kesehatan nasional. Dengan demikian, penggunaan teknologi machine learning seperti SVM dapat meningkatkan responsivitas dan efisiensi sistem kesehatan dalam menghadapi epidemi HIV/AIDS di Indonesia



Gambar 3. Time series plot

Pada Gambar 3, secara keseluruhan tren dari kolom [hiv_nasional] cenderung naik, meskipun memiliki beberapa penurunan ekstrim seperti pada periode 2015-2016 dan periode 2020-2022. Dimana kenaikan jumlah kasus terjadi pada setiap akhir tahun yang secara jelas menunjukkan bahwa adanya pola setiap tahunnya yang meningkat hampir setiap akhir tahun. Asumsi utama mengapa jumlah kasus HIV mengalami peningkatan di hampir setiap tahunnya dikarenakan peningkatan layanan tes dan tes yang dilakukan pada setiap akhir tahun. Asumsi ini dibuktikan pada Gambar 4.



Gambar 4. Matriks korelasi

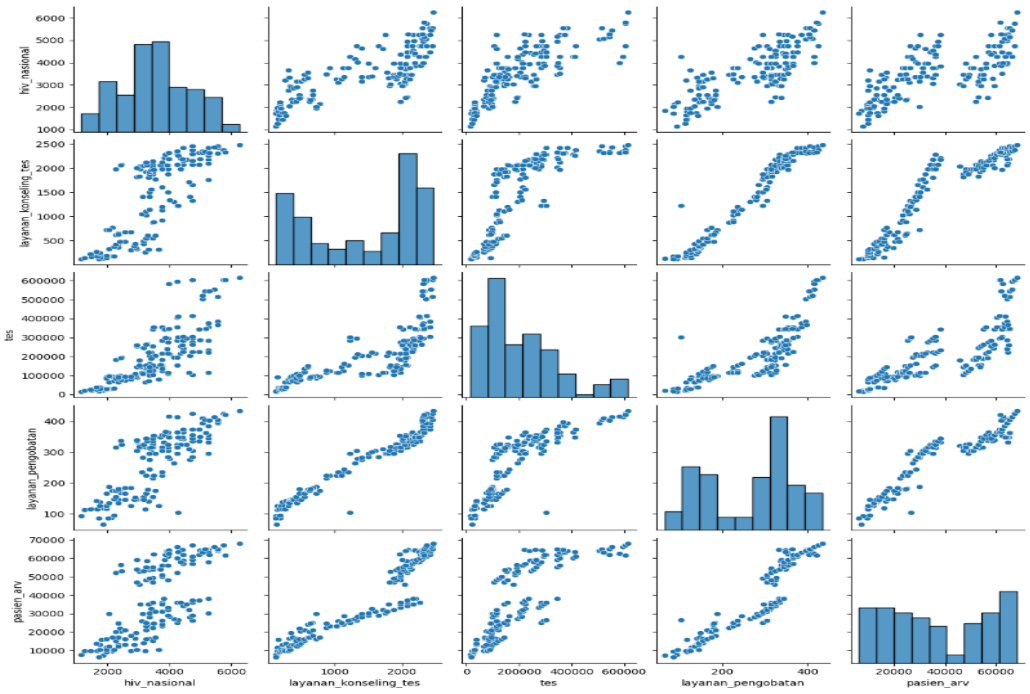
Nilai koefisien korelasi berkisar antara -1 hingga 1, di mana nilai -1 menunjukkan korelasi negatif sempurna, 0 menunjukkan tidak ada korelasi, dan 1 menunjukkan korelasi positif sempurna.

a. 1 = Korelasi Sempurna Positif
   Jika nilai korelasi antara dua variabel adalah 1, berarti keduanya bergerak searah secara sempurna. Artinya, setiap kali variabel X naik, variabel Y juga naik dengan proporsi yang tetap. Hubungan ini sangat kuat dan tidak ada penyimpangan dari pola linear positif
b. 0 = Tidak Ada Korelasi
   Nilai korelasi 0 menunjukkan bahwa tidak ada hubungan linear antara dua variabel. Artinya, perubahan pada variabel X tidak bisa digunakan untuk memprediksi perubahan pada variabel Y. Namun, ini tidak berarti tidak ada hubungan sama sekali—hanya tidak ada hubungan linear. Bisa saja ada hubungan non-linear.
c. -1 = Korelasi Sempurna Negatif

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

Jika nilai korelasi adalah -1, berarti kedua variabel memiliki hubungan linear yang berlawanan secara sempurna. Artinya, ketika variabel X naik, variabel Y akan turun dengan proporsi yang tetap. Ini menunjukkan hubungan negatif yang sangat kuat dan tanpa penyimpangan

Dari Gambar 4, terlihat bahwa variabel hiv_nasional memiliki korelasi cukup kuat terhadap keempat fitur eksternal lainnya, terutama dengan layanan_pengobatan (0.80) dan layanan_konseling_tes (0.79). Ini menunjukkan bahwa peningkatan layanan pengobatan dan konseling cenderung diikuti oleh peningkatan jumlah kasus HIV yang tercatat, yang masuk akal karena layanan tersebut bisa mendorong deteksi lebih banyak kasus. Hubungan antar fitur eksternal juga sangat tinggi, terutama antara layanan_konseling_tes dan layanan_pengobatan (0.98), serta dengan pasien_arv (0.93). Korelasi yang sangat tinggi ini bisa mengindikasikan adanya multikolinearitas. Artinya, fitur-fitur tersebut mungkin menyampaikan informasi yang sangat mirip. Dalam konteks model machine learning seperti SVM, hal ini bisa menyebabkan overfitting atau membingungkan model dalam mengestimasi kontribusi masing-masing fitur terhadap target. Secara keseluruhan, korelasi positif yang kuat antar fitur ini menunjukkan bahwa berbagai aspek layanan HIV saling terkait erat. Terlihat pada Gambar 4 tidak ada variabel dengan korelasi negatif. Hal ini disebabkan oleh tren epidemi yang makin tahun makin naik dikarenakan semua variabel yang diteliti (Gambar 2) memiliki sifat yang mengikuti tren epidemi itu sendiri. Seperti jumlah layanan konseling dan tes yang menentukan berapa jumlah pasien baru yang ditemukan, semakin banyak cakupan layanan tes dan konseling semakin banyak juga pasien batu yang akan ditemukan.



Gambar 5. Pair plot

Pada Gambar 5, pair plot ini menampilkan hubungan antar variabel numerik dalam dataset HIV nasional secara visual melalui kombinasi histogram dan scatter plot. Pada diagonal terdapat histogram distribusi masing-masing variabel, sedangkan di luar diagonal terlihat scatter plot yang menunjukkan hubungan antar pasangan variabel.

Distribusi hiv_nasional cenderung sedikit condong ke kanan (right-skewed), menandakan ada beberapa bulan dengan jumlah kasus sangat tinggi dibandingkan mayoritas lainnya. Scatter plot antara hiv_nasional dengan fitur seperti layanan_pengobatan, layanan_konseling_tes, dan pasien_arv menunjukkan hubungan positif yang cukup kuat dan linier, mendukung temuan korelasi sebelumnya. Ini menunjukkan bahwa semakin banyak layanan atau pasien yang ditangani, jumlah kasus yang tercatat juga meningkat.

Namun, terlihat juga bahwa beberapa fitur seperti tes memiliki sebaran data yang sangat lebar dan mengandung outlier yang signifikan, yang berpotensi mempengaruhi stabilitas model. Hubungan antar layanan_konseling_tes, layanan_pengobatan, dan pasien_arv tampak sangat linear dan rapat, mengindikasikan potensi multikolinearitas.

3.2.2 Implementasi/Pengujian (bila ada)

a. Data Preprocessing
Rekayasa fitur dilakukan dengan persiapan matriks X dan y dimana X berisi fitur, dan y sebagai variabel target. Adapun fitur yang digunakan di penelitian ini adalah X = [bulan (sebagai time based), layanan_konseling_tes, tes, layanan_pengobatan, pasien_arv (sebagai fitur numerik)] fitur ini dipilih berdasarkan asumsi bahwa variabel variabel ini memiliki ikatan yang berpotensi untuk mempengaruhi variabel target dan mampu memberikan gambaran kondisi secara real time pada waktu tertentu. Untuk variabel target sendiri adalah y = [hiv_nasional].

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

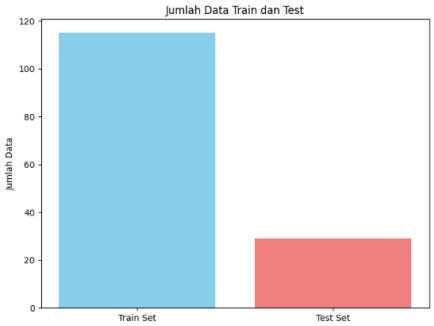
Tabel 1. Table matriks

Matriks X	Matriks y
Bulan	Hiv_nasional
Layanan_konseling_tes	
Tes	
Layanan_pengobatan	
Pasien_arv	

Tabel 1 menjabarkan seluruh matriks fitur (X) dan target (y) yang digunakan dalam model. Kolom 'Bulan' dan fitur layanan seperti 'layanan_konseling_tes', 'tes', 'layanan_pengobatan', dan 'pasien_arv' menjadi variabel input, sedangkan 'hiv_nasional' adalah variabel target yang diprediksi. Tabel ini menegaskan struktur data yang dipakai untuk pelatihan model SVR.

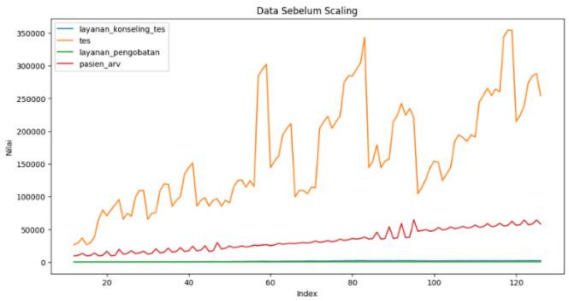
b. Data spliting dan scaling

Dari keseluruhan data, 80% akan dialokasikan sebagai training set dan 20% akan dialokasikan sebagai testing set. Dimana berarti dari total 156 baris, 125 bertindak sebagai training set dan 31 sebagai testing set.



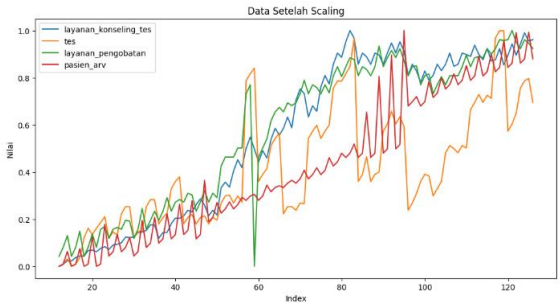
Gambar 6. Jumlah data train dan test

Pada Gambar 6 memperlihatkan pembagian dataset menjadi dua bagian: training set (80% data) dan testing set (20% data). Dari total 156 baris data, 125 baris digunakan untuk melatih model, sedangkan 31 baris sisanya dipakai untuk menguji keakuratan model. Visualisasi ini membantu memastikan pembagian data sudah proporsional dan memperlihatkan ukuran masing-masing subset.



Gambar 7. Data sebelum scalling

Pada Gambar 7 Menampilkan distribusi nilai fitur sebelum dilakukan normalisasi atau scaling. Terlihat ketimpangan nilai antar fitur, dimana kolom 'tes' memiliki nilai jauh lebih tinggi dibandingkan fitur lain. Ketimpangan ini bisa menyebabkan model SVM kesulitan dalam memproses data, sehingga perlu dilakukan scaling untuk menyamakan skala fitur.



Gambar 8. Data setelah scalling

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

Pada Gambar 8 menunjukkan hasil setelah proses normalisasi menggunakan MinMaxScaler, gambar ini menunjukkan distribusi nilai fitur yang sudah distandarisasi ke rentang nilai yang lebih seragam (biasanya 0-1). Terlihat jelas bahwa proses scaling membantu menyamakan nilai yang timpang jauh (**Gambar 7**) menjadi lebih seragam. Menunjukkan scaling ini sangat penting agar model SVM dapat bekerja secara optimal dan tidak bias terhadap fitur dengan skala besar.

c. Modelling

Model yang digunakan adalah SVR dengan epsilon-insensitive loss, dan RBF (Radial Basis Function) sebagai kernel. Adapun parameter lengkap yang akan di tuning:

Tabel 2. Nilai Parameter

Parameter	Nilai
C	{0.1 , 1 , 10 , 100 , 1000}
Epsilon	{0.1}
Gamma	{scale, auto, 0.001, 0.01, 0.1}

Tabel 2 mencantumkan rentang nilai parameter yang diuji dalam proses tuning model SVR, termasuk nilai C (regularisasi), epsilon (margin toleransi kesalahan), dan gamma (parameter kernel RBF). Parameter ini sangat menentukan performa akhir model, sehingga perlu diuji dengan metode grid search untuk mendapatkan kombinasi terbaik. Selanjutnya adalah implementasi Grid Search dengan Time Series Cross Validation untuk mencari kombinasi parameter dengan 'score' yang terbaik.

Tabel 3. Parameter Gridsearch

Grid Search	
Model	SVR
Parameter	{C, Epsilon, Gamma, Kernel}
Cross Validation	Time Series Cross Validation (tscv = 5)
Verbose	10

Tabel 3 menjelaskan setup dari proses grid search yang digunakan untuk mencari kombinasi parameter terbaik model SVR. Tabel memuat nama model (SVR), parameter yang diuji, jenis validasi (Time Series Cross Validation dengan 5 fold), dan tingkat verbosity selama proses pencarian parameter. Ini menunjukkan bagaimana optimasi model dilakukan secara sistematis. GridSearch menghasilkan 25 kandidat kombinasi parameter dan 125 fits, dengan hasil :

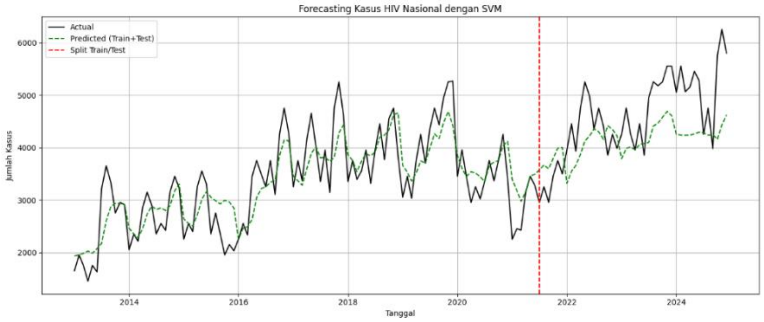
Tabel 4. Hasil gridsearch

params	mean_test_score	std_test_score	rank_test_score
{'C': 1000, 'epsilon': 0.1, 'gamma': 0.1, 'kernel': 'rbf'}	-0,511791162	1,461350592	1
{'C': 1000, 'epsilon': 0.1, 'gamma': 'auto', 'kernel': 'rbf'}	-0,683832912	1,837072829	2
{'C': 1000, 'epsilon': 0.1, 'gamma': 'scale', 'kernel': 'rbf'}	-1,159115183	2,682713837	3
{'C': 1000, 'epsilon': 0.1, 'gamma': 0.01, 'kernel': 'rbf'}	-1,429824266	2,549487534	4
{'C': 100, 'epsilon': 0.1, 'gamma': 'scale', 'kernel': 'rbf'}	-1,497460429	2,500036497	5

Hasil pencarian parameter terbaik yang sudah ditentukan (Tabel 3) dengan Grid Search, didapatkan parameter optimal dengan skor evaluasi (mean_test_score) sebesar -0,511791162. Yaitu kombinasi {'C': 1000, 'epsilon': 0.1, 'gamma': 0.1, 'kernel': 'rbf'} seperti yang dapat di lihat pada Tabel 4.

d. Hasil peramalan

Pada tahap ini, dilakukan evaluasi hasil peramalan yang dihasilkan oleh model Support Vector Regression (SVR) pada data pengujian (test set). Tabel di bawah ini menunjukkan perbandingan antara nilai aktual (y_test) dan nilai prediksi (y_pred) untuk periode yang berbeda (Tabel 5). Peramalan dilakukan untuk memprediksi jumlah kasus HIV/AIDS pada bulan-bulan mendatang berdasarkan data historis yang tersedia.



Gambar 9. Peramalan kasus HIV nasional dengan SVM

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

Pada plot hasil prediksi model Support Vector Regression (SVR) dapat terlihat ada beberapa penurunan nilai pada nilai dibandingkan dengan data aktual kasus HIV nasional pada data testing. Grafik ini memperlihatkan tren prediksi model yang cukup mengikuti pola data asli, dengan nilai prediksi yang mendekati nilai aktual pada sebagian besar periode, meskipun terdapat beberapa selisih kecil. Visualisasi ini menggambarkan performa model secara nyata dalam memprediksi jumlah kasus HIV. Untuk penjabaran hasil peramalan secara lengkap (hasil pada test set):

Tabel 5. Hasil peramalan

Indeks	y_test	y_pred
2022-08-01	4755	4303.754957
2022-09-01	4414	4161.611578
2022-10-01	3855	4418.723459
2022-11-01	4255	4346.761827
2022-12-01	3987	4227.691577
2023-01-01	4255	3789.991715
2023-02-01	4755	3983.056224
2023-03-01	4269	4017.571253
2023-04-01	3955	3959.852674
2023-05-01	4455	4060.503786
2023-06-01	3857	4077.447328
2023-07-01	4955	4098.697262
2023-08-01	5255	4411.588812
2023-09-01	5179	4465.745173
2023-10-01	5255	4579.943807
2023-11-01	5555	4689.591172
2023-12-01	5554	4611.132277
2024-01-01	5055	4256.005789
2024-02-01	5555	4239.118506
2024-03-01	5065	4237.457649
2024-04-01	5155	4237.782735
2024-05-01	5455	4262.887062
2024-06-01	5279	4295.401664
2024-07-01	4255	4285.060544
2024-08-01	4755	4236.791825
2024-09-01	3981	4250.089274
2024-10-01	5755	4160.186091
2024-11-01	6255	4433.702923
2024-12-01	5802	4626.157170

Terlihat terdapat perbedaan yang cukup signifikan pada nilai aktual (y_test) dengan nilai prediksi (y_pred) dari model SVR pada data testing pada beberapa periode waktu. Tabel 4 ini memberikan rincian numerik tentang ketepatan prediksi, menunjukkan bahwa meskipun ada beberapa perbedaan, model mampu mendekati nilai sebenarnya dengan cukup baik.

Penjelasan Hasil:

1. y_test (Nilai Aktual): Merupakan nilai aktual yang tercatat pada data untuk setiap periode waktu. Ini adalah data yang digunakan sebagai referensi untuk membandingkan akurasi hasil peramalan.
2. y_pred (Nilai Prediksi): Merupakan hasil prediksi yang dihasilkan oleh model SVR untuk periode yang sama. Nilai ini menunjukkan estimasi jumlah kasus HIV/AIDS yang diprediksi oleh model berdasarkan data pelatihan.

Dari hasil peramalan di atas, dapat dilihat bahwa nilai prediksi cukup mendekati nilai aktual untuk sebagian besar periode, meskipun terdapat beberapa perbedaan antara y_test dan y_pred. Misalnya, pada bulan 2022-08-01, nilai aktual adalah 4755, sementara nilai prediksi adalah 4303.75. Perbedaan ini mengindikasikan adanya ketidaktepatan kecil dalam prediksi, yang dapat terjadi karena berbagai faktor yang tidak tercakup dalam model, seperti intervensi medis atau perubahan perilaku Masyarakat hal ini menjelaskan perbedaan garis plot hasil (Gambar 9).

Secara umum, model ini menunjukkan kemampuan yang baik dalam meramalkan jumlah kasus HIV/AIDS di masa depan dengan tingkat kesalahan yang terkontrol, yang ditunjukkan oleh perbedaan yang relatif kecil antara nilai prediksi dan nilai aktual, meskipun ada beberapa fluktuasi yang perlu diperhatikan.

3.2 Pembahasan

Evaluasi model dilakukan untuk menilai sejauh mana model yang dikembangkan dapat memprediksi dengan akurat berdasarkan data yang tersedia. Dalam penelitian ini, model yang digunakan adalah Support Vector Regression (SVR),

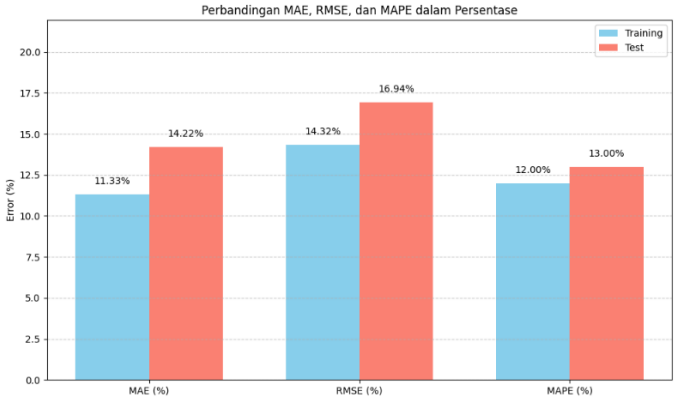
Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

yang diterapkan pada data time series untuk meramalkan jumlah kasus HIV/AIDS. Evaluasi model menggunakan tiga metrik utama, yaitu Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan Mean Absolute Percentage Error (MAPE).MAE mengukur rata-rata perbedaan absolut antara nilai prediksi dan nilai aktual, yang memberikan gambaran mengenai seberapa besar kesalahan prediksi secara umum tanpa memperhitungkan arah kesalahan. RMSE mengukur akar kuadrat dari rata-rata kesalahan kuadrat, memberikan penalti yang lebih besar terhadap kesalahan yang lebih besar dan menggambarkan ketepatan prediksi secara keseluruhan. MAPE mengukur rata-rata persentase kesalahan antara nilai prediksi dan nilai aktual, memberikan gambaran mengenai kesalahan relatif yang terjadi dalam prediksi model. MAPE juga memberikan interpretasi yang lebih mudah dipahami karena disajikan dalam bentuk persentase.

Tabel 6. Metrik evaluasi

Training/Testing	MAE	RMSE	MAPE
Training	383.69	485.36	0.12 (12%)
Testing	691.34	823.11	0.13 (13%)

Tabel 6 menyajikan nilai tiga metrik evaluasi model pada data training dan testing: MAE, RMSE, dan MAPE. Nilai-nilai ini menunjukkan tingkat kesalahan rata-rata, kesalahan kuadrat, dan persentase kesalahan relatif prediksi model. Metrik ini digunakan untuk menilai seberapa akurat dan stabil model dalam memprediksi kasus HIV. Terlihat bahwa meskipun kesalahan sedikit meningkat pada data testing, performa model masih berada pada tingkat yang dapat diterima.



Gambar 10. Perbandingan MAE, RMSE, dan MAPE dalam persentase.

Gambar 10 merupakan visualisasi perbandingan MAE, RMSE, MAPE yang menunjukan perbandingan performa SVR Training Set: Pada data pelatihan, model SVR menunjukkan nilai MAE sebesar 383.69, RMSE sebesar 485.36, dan MAPE sebesar 12%. MAPE yang rendah ini menunjukkan bahwa model dapat memprediksi nilai secara akurat dengan kesalahan relatif yang tidak terlalu besar. Testing Set: Pada data pengujian, model SVR memberikan MAE sebesar 691.34, RMSE sebesar 823.11, dan MAPE sebesar 13%. Meskipun terjadi sedikit peningkatan kesalahan pada data pengujian, nilai MAPE yang masih berada di angka 13% (Tabel 6)menunjukkan bahwa model tetap memiliki kinerja yang baik dan mampu memberikan prediksi yang dapat diandalkan.

Secara keseluruhan, model SVR menunjukkan performa yang cukup baik baik pada data pelatihan (training set) maupun data pengujian (testing set). Metrik evaluasi menunjukkan bahwa model ini dapat memberikan prediksi dengan akurasi yang baik, dengan MAPE yang berada di kisaran 12-13% dan metrik lainnya (MAE dan RMSE) berada pada rentang yang dapat diterima (11-17%). Ini menunjukkan bahwa meskipun terdapat sedikit perbedaan dalam kinerja antara data pelatihan dan pengujian, model SVR tetap efektif dalam memprediksi jumlah kasus HIV/AIDS berdasarkan data yang tersedia.

4. KESIMPULAN

Berdasarkan hasil analisis dan penerapan model Support Vector Machine (SVM) dalam memprediksi jumlah kasus HIV/AIDS di Indonesia, dapat disimpulkan bahwa model SVM memberikan performa yang baik dan efektif dalam menghadapi data yang bersifat nonlinier dan dinamis. Penggunaan data time series dari periode 2012 hingga 2024 memungkinkan model ini untuk menangkap pola tren penyebaran HIV/AIDS dengan tingkat akurasi yang lebih tinggi dibandingkan metode tradisional seperti regresi linier dan ARIMA. Variabel-variabel layanan kesehatan seperti layanan konseling, tes HIV, dan pengobatan antiretroviral terbukti berpengaruh signifikan terhadap prediksi jumlah kasus, yang menunjukkan hubungan kompleks antar faktor epidemiologi yang berhasil dimodelkan oleh SVM. Model ini juga mampu menangani keterbatasan jumlah data serta mengurangi risiko overfitting, sehingga menghasilkan prediksi yang stabil dan andal. Hasil evaluasi menggunakan metrik MAE, RMSE, dan MAPE mengindikasikan bahwa model SVR dengan kernel RBF memiliki tingkat kesalahan yang rendah dan dapat diandalkan, baik pada data

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 465−475
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7365

pelatihan maupun pengujian. Dengan demikian, model ini berpotensi menjadi alat bantu yang efektif bagi pengambil kebijakan dalam merancang strategi pencegahan dan pengendalian HIV/AIDS yang berbasis data. Penelitian ini membuka peluang untuk pengembangan lebih lanjut, seperti integrasi data eksternal yang lebih beragam, peningkatan fitur, serta implementasi sistem prediksi yang terintegrasi dalam monitoring kesehatan nasional. Dengan demikian, penggunaan teknologi machine learning seperti SVM dapat meningkatkan responsivitas dan efisiensi sistem kesehatan dalam menghadapi epidemi HIV/AIDS di Indonesia

REFERENCES

[1] E. Widiastuti, A. Ika Fibriana, "Kejadian HIV/AIDS di Kota Semarang Tahun 2021," *HIGEIA (Journal of Public Health Research and Development)*, vol 6, no 4, 2022, doi: 10.15294/higeia/v6i4/57060.
[2] H. Wiguna, Y. Nugraha, F. Rizka R, A. Andika, J. I. Kanggrawan, and A. L. Suherman, "Kebijakan Berbasis Data: Analisis dan Prediksi Penyebaran COVID-19 di Jakarta dengan Metode Autoregressive Integrated Moving Average (ARIMA)," *Jurnal Sistem Cerdas*, vol. 3, no. 2, pp. 74–83, Aug. 2020, doi: 10.37396/jsc.v3i2.76.
[3] W. Gunawan, Y. Devianto, and A. P. Sari, "Imbalanced Data NearMiss for Comparison of SVM and Naive Bayes Algorithms," *Computer Engineering and Applications Journal*, vol. 13, no. 03, pp. 34–43, Oct. 2024, doi: 10.18495/comengapp.v13i03.485.
[4] N. Rusmilawati and P. T. Prasetyaningrum, "Penerapan Data Mining Dalam Prediksi Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah Menggunaka Metode Linier Regression," *Journal of Information System and Artificial Intelligence (JISAI)*, vol. 1, no. 2, pp. 1–7, 2021, doi: https://doi.org/10.46338/jisai.v1i2.33.
[5] H. Maika, *Manajemen Epidemiologi*, 1st ed. Padang: GET PRESS INDONESIA, 2024. [Online]. Available: https://www.researchgate.net/publication/387512413
[6] M. L. Bovbjerg, *Foundations of Epidemiology*. Oregon State Open Educational Resources, 2020.
[7] W. Wahyuningsih and P. T. Prasetyaningrum, "Enhancing Sales Determination for Coffee Shop Packages through Associated Data Mining: Leveraging the FP-Growth Algorithm," *Journal of Information Systems and Informatics*, vol. 5, no. 2, pp. 758–770, May 2023, doi: 10.51519/journalisi.v5i2.500.
[8] M. Oktafani and P. T. Prasetyaningrum, "Implementasi Support Vector Machine Untuk Analisis Sentimen Komentar Aplikasi Tanda Tangan Digital," *Jurnal Sistem Informasi dan Bisnis Cerdas*, vol. 15, no. 1, Mar. 2022, doi: 10.33005/sibc.v15i1.2697.
[9] A. Kloska, A. Harmoza, S. M. Kloska, T. Marciniak, and I. Sadowska-Krawczenko, "Predicting preterm birth using machine learning methods," *Sci Rep*, vol. 15, no. 1, p. 5683, Feb. 2025, doi: 10.1038/s41598-025-89905-1.
[10] N. Shahira Pisal, S. Abdul-Rahman, M. Hanafiah, and S. I. Kamarudin, "Prediction Of Life Expectancy For Asian Population Using Machine Learning Algorithms," *Malaysian Journal of Computing*, vol. 7, no. 2, pp. 1150–1161, 2022, doi: 10.24191/mjoc.v7i2.18218.
[11] P. T. Prasetyaningrum, P. Purwanto, and A. F. Rochim, "Enhancing User Engagement in Mobile Banking Through Personalized Gamification: A Cognitive Evaluation Theory Approach," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 4, pp. 788–807, 2024, doi: 10.22266/IJIES2024.0831.60.
[12] S. U. Nisa, A. Mahmood, F. S. Ujager, and M. Malik, "HIV/AIDS predictive model using random forest based on socio-demographical, biological and behavioral data," *Egyptian Informatics Journal*, vol. 24, no. 1, pp. 107–115, Mar. 2023, doi: 10.1016/j.eij.2022.12.005.
[13] A. Mustika Rani and N. Hendrastuty, "Perbandingan Algoritma NBC Dan SVM Untuk Melakukan Analisis Sentimen Terhadap PP NO.82 Tahun 2021," *Technology and Science (BITS)*, vol. 6, no. 4, 2025, doi: 10.47065/bits.v6i4.6496.
[14] R. N. Azizah and U. Indahyanti, "Analysis of the Predicted Number of HIV/AIDS Spreads in Sidoarjo Regency using Multiple Linear Regression Method," *Advances in Cancer Science*, vol. 1, no. 1, p. 11, Apr. 2024, doi: 10.47134/acsc.v1i1.3.
[15] M. M. Da Costa, "Sistem Pakar Mendeteksi Penyakit Hiv/Aids Di Ntt Dengan Metode Support Vector Regression (SVR)," *Prosiding Seminar Nasional SEMMAU 2019*, vol. tidak tersedia, no. tidak tersedia, pp. 856–862, 2022, doi: http://dx.doi.org/10.33364/algoritma/v.19-2.1168.
[16] S. S. and M. Elkanzi, "A Review on the Role of Machine Learning in Predicting the Spread of Infectious Diseases," *Metaheuristic Optimization Review*, vol. 2, no. 1, pp. 14–27, 2024, doi: 10.54216/MOR.020102.
[17] D. Mesafint Belete, "Wrapper Based Feature Selection Techniques On EDHS-HIV/AIDS Dataset," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 8, pp. 2642–2657, 2020, doi: https://doi.org/10.31838/ejmcm.07.08.274.
[18] A. Abade *et al.*, "A comparative analysis of classical and machine learning methods for forecasting TB/HIV co-infection," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-69580-4.
[19] A. K. Sah *et al.*, "Role of Artificial Intelligence and Personalized Medicine in Enhancing HIV Management and Treatment Outcomes," *Life*, vol. 15, no. 5, p. 745, May 2025, doi: 10.3390/life15050745.
[20] D. Tang *et al.*, "Study on the prediction performance of AIDS monthly incidence in Xinjiang based on time series and deep learning models," *BMC Public Health*, vol. 25, no. 1, p. 780, Feb. 2025, doi: 10.1186/s12889-025-21982-3.
[21] M. Jordan, J. Kleinberg, and B. Schölkopf, *Pattern Recognition and Machine Learning*. Springer Science, 2006.