

Perbandingan Algoritma Linear Regression, Support Vector Regression, dan Artificial Neural Network untuk Prediksi Data Obat

Suci Maharani Putri*, Rice Novita, Mustakim, M Afdal

Fakultas Sains dan Teknologi, Prodi Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}11950321583@students.uin-suska.ac.id, ²rice-novita@uin-suska.ac.id, ³mustakim@uin-suska.ac.id,

⁴m.afdal@uin-suska.ac.id

Email Penulis Korespondensi: 11950321583@students.uin-suska.ac.id

Submitted: 15/05/2024; Accepted: 30/05/2024; Published: 23/06/2024

Abstrak—Regresi menjadi fokus penting di berbagai bidang yang bertujuan untuk membuat perkiraan nilai masa depan untuk membantu pengambilan keputusan dan membuat perencanaan strategis. Berbagai algoritma regresi memiliki kelebihan dan kekurangannya masing-masing dan performanya dapat berbeda sesuai dengan karakteristik data yang digunakan. Untuk itu analisis lebih lanjut diperlukan untuk mengetahui algoritma yang tepat yang menghasilkan solusi terbaik dari masalah yang akan diselesaikan. Penelitian ini membandingkan tiga algoritma regresi populer yaitu Linear Regression (LR), Support Vector Regression (SVR), dan Artificial Neural Network (ANN) untuk memprediksi data obat pada salah satu apotek di provinsi Riau. Hingga saat ini, pihak apotek masih belum memiliki metode yang tepat untuk memperkirakan kebutuhan obat setiap bulan dan masih dilakukan berdasarkan perkiraan. Hal tersebut membuat apotek sering mengalami kekurangan bahkan kelebihan stok yang dapat menyebabkan kerugian terlebih jika masa pakai obat sudah terlewat. Tiga jenis obat dipilih untuk menguji kemampuan algoritma yang diusulkan yaitu Amoxicilin, Antasida dan Paracetamol. Melalui hasil analisis dan perbandingan menunjukkan bahwa algoritma SVR lebih unggul dibandingkan algoritma lainnya pada ketiga jenis data obat jika konteks dan analisis berfokus pada metrik RMSE. Sementara itu jika konteks dan analisis berfokus pada metrik MAPE, maka algoritma ANN lebih unggul dibandingkan algoritma lainnya. Meskipun LR tidak lebih unggul pada metrik apapun, nyatanya ketiga algoritma (LR, SVR, dan ANN) memiliki nilai MAPE dibawah 10%, yang menunjukkan hasil prediksinya masih dalam kategori sangat akurat. Hal tersebut dibuktikan oleh hasil prediksi dari seluruh model yang diusulkan mampu mengikuti pola dan tren pada data aktual dengan baik.

Kata Kunci: Artificial Neural Network; Linear Regression; Support Vector Regression; Prediksi; Obat

Abstract—Regression is a crucial focus in various fields aiming to forecast future values to aid decision-making and strategic planning. Different regression algorithms have their advantages and disadvantages, and their performance can vary depending on the data characteristics. Therefore, further analysis is needed to identify the appropriate algorithm that provides the best solution for the problem at hand. This study compares three popular regression algorithms: Linear Regression (LR), Support Vector Regression (SVR), and Artificial Neural Network (ANN) to predict drug data at a pharmacy in Riau province. Currently, the pharmacy lacks an accurate method for estimating monthly drug needs, relying instead on rough estimates. This often results in either shortages or overstock, leading to losses, especially if the drugs expire. Three types of drugs, namely Amoxicillin, Antacids, and Paracetamol were selected to test the proposed algorithms. The analysis and comparison show that the SVR algorithm outperforms the others on all three drug types when focusing on the RMSE metric. However, when the focus is on the MAPE metric, the ANN algorithm proves to be superior. Although LR does not excel in any metric, all three algorithms (LR, SVR, and ANN) have MAPE values below 10%, indicating highly accurate predictions. This accuracy is evidenced by the prediction results of all proposed models, which effectively follow the patterns and trends in the actual data.

Keywords: Artificial Neural Network; Linear Regression; Support Vector Regression; Prediction; Drugs

1. PENDAHULUAN

Machine learning (ML) merupakan ilmu yang mempelajari bagaimana komputer dapat belajar tanpa diprogram secara eksplisit menggunakan data historis sehingga dapat membantu pengambilan keputusan [1]. Terdapat banyak tugas yang dapat dilakukan oleh ML, salah satunya adalah regresi. Regresi adalah salah satu teknik ML yang digunakan untuk memprediksi nilai numerik dari variabel terikat berdasarkan variabel prediktor [2]. Penerapan regresi dapat ditemui dalam berbagai bidang, termasuk pada penjualan suatu bisnis. Tujuan dari regresi adalah membuat perkiraan nilai di masa depan untuk membantu pengambilan keputusan dan perencanaan yang strategis sehingga dapat meminimalkan risiko yang mungkin terjadi.

Terdapat banyak algoritma ML yang dapat diimplementasikan untuk melakukan regresi. Beberapa yang populer yaitu *Linear Regression* (LR), *Support Vector Regression* (SVR), dan *Artificial Neural Network* (ANN). Algoritma LR merupakan pemodelan berbasis statistik untuk melakukan prediksi yang melibatkan dua atau lebih variabel yaitu variabel pemberi pengaruh dan variabel terpengaruh [3], [4]. Selanjutnya, SVR merupakan penerapan SVM pada kasus regresi [5] dengan menemukan fungsi sebagai *hyperplane* dengan kesalahan sekecil mungkin [6]. Kemudian ANN merupakan algoritma yang terinspirasi dari struktur jaringan saraf biologis manusia [7]. Algoritma ini memiliki arsitektur yang terdiri dari lapisan masukan (*input layer*), satu atau lebih lapisan tersembunyi (*hidden layer*), dan lapisan keluaran (*output layer*).

Setiap algoritma tersebut tentu memiliki kelebihan dan kekurangannya masing-masing. Algoritma LR sangat baik untuk melakukan prediksi jangka pendek [8]. Tidak hanya itu, algoritma LR termasuk sederhana sehingga mudah diimplementasikan. Namun LR memiliki kelemahan, yaitu sangat bergantung pada asumsi bahwa hubungan antara variabel adalah linear, sehingga tidak mampu menangkap dan memodelkan data non-linear [8]. Selanjutnya algoritma SVR efektif memodelkan data berdimensi tinggi, serta dapat menangani data linear dan non-linear dengan fungsi

kernel [9]. Namun SVR kekurangannya adalah terdapat banyak parameter yang kompleks, sehingga sulit untuk memilih parameter yang tepat [10]. Kemudian ANN memiliki kelebihan dapat menangani masalah prediksi non-linier yang kompleks [11], [12]. Namun kekurangannya, ANN memerlukan data dan waktu pelatihan yang lebih banyak. Selain itu, algoritma ANN memiliki konfigurasi parameter yang kompleks sehingga sulit untuk menemukan parameter yang tepat, terutama pada jumlah neuron dan lapisannya [7].

Terdapat beberapa penelitian sebelumnya yang telah menggunakan algoritma tersebut untuk menangani kasus regresi. Penelitian yang dilakukan oleh S. Agnesti et. al pada tahun 2023 membandingkan algoritma SVR dan *Triple Exponential Smoothing* (TES) untuk prediksi jumlah pemakaian pada dua jenis obat menghasilkan SVR sebagai yang terbaik dengan nilai MAPE berturut-turut yaitu 10.31% dan 9.38% [13]. Meskipun SVR memberikan hasil yang baik, namun penelitian ini masih memiliki kelemahan, yaitu algoritma SVR yang diimplementasikan hanya menggunakan satu jenis kernel saja, yaitu kernel RBF tanpa menguji kernel lainnya. Penelitian selanjutnya dilakukan oleh S. Pratista et. al pada tahun 2023 membandingkan SVR dan *Single Moving Average* (SMA) untuk prediksi jumlah pemakaian pada dua jenis obat menghasilkan SMA sebagai model terbaik dengan MAPE sebesar 4.29% dan 4.10% [14]. Meskipun SMA terbukti lebih baik dibandingkan SVR, namun penggunaan algoritma SVR hanya diuji menggunakan satu kernel yaitu RBF, seperti pada penelitian S. Agnesti et. al pada tahun 2023 [13]. Selain itu, kedua penelitian tersebut masih dapat dikembangkan lebih lanjut dengan menerapkan metode yang lebih canggih, seperti algoritma berbasis *neural network*.

Selanjutnya, penelitian yang dilakukan oleh R. Novita et. al pada tahun 2022 menggunakan algoritma LR pada sistem prediksi untuk menentukan jumlah pemesanan obat menghasilkan nilai MAPE sebesar 12.42% [15]. Meskipun hasil yang didapatkan memiliki kategori prediksi yang akurat, namun penelitian ini hanya menguji satu algoritma saja yaitu LR, yang memiliki batasan tidak dapat memodelkan data non-linear. Kemudian, penelitian yang dilakukan oleh Z. Jannah & K. Budayawan pada tahun 2024 melakukan prediksi kebutuhan obat menggunakan algoritma ANN menghasilkan MAPE sebesar 6.817% [12]. Hasil prediksi yang didapatkan memiliki akurasi yang sangat akurat, namun model hanya diuji pada satu jenis obat dan hanya menggunakan parameter *learning rate* 0.1 dengan 1000 iterasi tanpa melakukan pengujian *hyperparameter* lebih lanjut, seperti penggunaan nilai *learning rate* yang berbeda ataupun *optimizer* untuk mendapatkan model yang lebih optimal. Penelitian lainnya dilakukan oleh N. E. Warestika et. al pada tahun 2022 melakukan peramalan stok obat dengan ANN menghasilkan nilai MSE sebesar 67.72 dan RMSE sebesar 8.22 [16]. Sayangnya, penelitian ini tidak menuliskan informasi secara rinci tentang *hyperparameter* yang digunakan pada penelitiannya.

Berdasarkan penelitian sebelumnya, terlihat bahwa tidak semua algoritma dapat dengan baik memahami pola dan tren pada data yang dimodelkan sehingga menghasilkan kemampuan prediksi yang berbeda-beda. Untuk itu, penelitian ini melakukan perbandingan tiga algoritma regresi ML, yaitu LR, SVR dan ANN untuk mengembangkan model prediksi yang efektif dan akurat. Perbedaan penelitian ini dengan penelitian sebelumnya [12]–[16] adalah pada penelitian ini mengusulkan algoritma regresi yang belum pernah dilakukan dan dibandingkan oleh penelitian sebelumnya dalam satu penelitian dengan kasus yang sama. Selain algoritma, perbedaan lainnya adalah pada penelitian ini melakukan eksperimen yang berbeda dan mendalam untuk mengisi gap pada penelitian sebelumnya sehingga membuat penelitian ini unik dan memiliki kebaharuan.

Lebih lanjut, penelitian ini menggunakan tiga jenis kernel pada algoritma SVR (linear, RBF dan polynomial) dan menguji algoritma yang lebih canggih berbasis *neural network* yaitu ANN, sehingga dapat memberikan hasil pengujian yang lebih mendalam dibandingkan penelitian S. Agnesti et. al pada tahun 2023 dan S. Pratista et. al pada tahun 2023, yang hanya menggunakan satu kernel pada SVR [13], [14]. Selanjutnya, penggunaan ANN dan SVR dengan beberapa kernel mampu memodelkan data non-linear, sehingga dapat melengkapi penelitian R. Novita et. al pada tahun 2022 yang hanya menggunakan LR untuk melakukan prediksi [15]. Kemudian, penelitian Z. Jannah & K. Budayawan pada tahun 2024 dan N. E. Warestika et. al tahun 2022 tidak melakukan pengujian parameter ANN secara mendalam [12], [16], namun penelitian ini melakukan pengujian parameter untuk mendapatkan model yang optimal, terutama pada *learning rate* dan *optimizer*. Selain itu, penelitian yang dilakukan oleh Z. Jannah & K. Budayawan tahun 2024, S. Agnesti et. al pada tahun 2023, dan S. Pratista et. al tahun 2023 hanya menggunakan satu atau dua jenis obat [12]–[14], sedangkan penelitian ini menggunakan tiga jenis obat untuk menguji algoritma yang diusulkan.

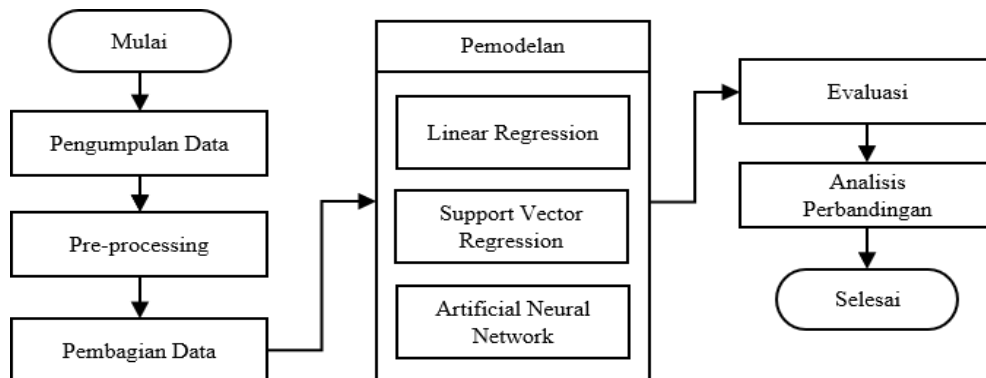
Tidak hanya itu, dataset yang digunakan penelitian ini juga berbeda dengan penelitian yang telah ada, yaitu data penjualan dan pembelian obat di salah satu apotek yang berlokasi di Air Tiris, Kecamatan Kampar, Provinsi Riau. Apotek ini dipilih karena memiliki jumlah permintaan obat yang cukup tinggi, sehingga memerlukan analisis yang tepat untuk memperkirakan jumlah kebutuhannya. Banyaknya kebutuhan obat membuat pihak apotek sering kekurangan stok. Melalui data dan informasi yang di dapatkan dari pihak apotek, pemesanan obat yang dilakukan dapat mencapai 100 box. Selain itu, apotek ini juga pernah mengalami kelebihan persediaan, sehingga menimbulkan kerugian bagi pihak apotek, terlebih jika masa pakai obat sudah terlewat. Hal tersebut pernah terjadi pada tahun 2022 yang mengalami kelebihan obat mencapai 1000 pcs terutama pada kategori obat penyakit khusus.

Hingga saat ini, pihak apotek masih belum memiliki metode yang tepat untuk memperkirakan kebutuhan obat dan masih melakukan pemesanan berdasarkan perkiraan. Untuk menjawab permasalahan tersebut, maka penelitian ini akan melakukan prediksi penjualan obat menggunakan tiga algoritma regresi ML yang populer, yaitu LR, SVR, dan ANN untuk dapat memperkirakan stok obat sesuai dengan kebutuhan. Penelitian ini diharapkan dapat menjadi solusi bagi apotek sehingga dapat mengurangi resiko terjadinya kerugian akibat kekurangan ataupun kelebihan stok obat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini melakukan perbandingan algoritma regresi berbasis *machine learning* (ML) pada data obat apotek. Seluruh tahapan penelitian digambarkan melalui diagram alir yang dapat dilihat pada gambar 1.



Gambar 1. Metodologi Penelitian

Gambar 1 secara garis besar menjelaskan seluruh tahapan penelitian ini dimulai dari pengumpulan data secara langsung pada pihak apotek, melakukan *pre-processing* untuk memperbaiki format data, pembagian data menggunakan teknik *hold-out validation*, pemodelan menggunakan tiga algoritma ML, evaluasi model berdasarkan nilai kesalahan (*error*) prediksi terhadap data aktual, dan analisis perbandingan untuk mendapatkan model terbaik pada kasus prediksi data obat.

2.2 Pengumpulan Data

Tahapan pengumpulan data dilakukan dengan observasi dan pengambilan secara langsung ke pihak Apotek. Rentang waktu pengambilan data dimulai dari bulan Januari 2021 hingga Desember 2023 dengan total 36 baris data. Terdapat tiga data jenis obat yang digunakan pada penelitian ini yaitu Amoxicilin, Antasida, dan Paracetamol. Pemilihan tersebut didasari oleh ketiga jenis obat tersebut sering mengalami kehabisan stok di apotek.

2.3 Pre-processing

Tahapan *pre-processing* bertujuan untuk menyesuaikan format data sehingga siap untuk dianalisis. Beberapa proses yang dilakukan pada tahap ini yaitu *cleaning*, ekstraksi fitur, dan normalisasi. Proses *cleaning* akan membersihkan data dari fitur yang tidak relevan. Kemudian ekstraksi fitur dilakukan untuk membuat *sequence* pada data sehingga membentuk fitur temporal. Terakhir yaitu normalisasi yang akan mengubah rentang nilai pada data menjadi skala 0 hingga 1 menggunakan teknik *min-max normalization*.

2.4 Pembagian Data

Teknik pembagian data yang digunakan pada penelitian ini adalah *hold-out validation*. Teknik ini sederhana namun populer dan terbukti efektif. *Hold-out* akan membagi dataset menjadi dua bagian, yaitu data pelatihan dan data pengujian [17]. Proporsi pembagian data yang digunakan pada penelitian ini yaitu 80%:20%, yang artinya 80% dari total data akan dialokasikan sebagai data pelatihan dan 20% sebagai data pengujian.

2.5 Pemodelan

2.5.1 Linear Regression

Linear regression (LR) merupakan salah satu pemodelan statistik untuk melakukan prediksi yang melibatkan variabel dependen (Y) dan independen (X) [3], [4]. Model *simple linear regression* melibatkan setidaknya satu variabel independen, sedangkan *multiple linear regression* melibatkan dua atau lebih variabel independen [1], [4]. Algoritma ini sangat baik untuk melakukan prediksi jangka pendek [8]. Pada implementasinya, LR akan menghasilkan persamaan garis lurus yang nantinya akan digunakan sebagai pedoman untuk memprediksi nilai pada waktu yang mendatang [3]. Hal inilah yang membuat algoritma ini sederhana dan mudah diimplementasikan. Namun algoritma LR memiliki kelemahan, yaitu sangat bergantung pada asumsi bahwa hubungan antara variabel adalah linear, sehingga algoritma ini tidak mampu menangkap dan memodelkan data nonlinear [8].

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i \quad (1)$$

2.5.2 Support Vector Regression

Support Vector Regression (SVR) merupakan penerapan *support vector machine* (SVM) pada kasus regresi yang dimodifikasi dengan menambahkan variabel baru yaitu epsilon [5]. Cara kerja SVR yaitu dengan menemukan fungsi

sebagai hyperplane berupa fungsi regresi yang mencocokkan seluruh data masukan dengan kesalahan sekecil mungkin [6]. Algoritma ini memiliki kelebihan yaitu dapat digunakan pada kasus linear maupun non-linear dengan fungsi kernel [9], [18]. Selain itu, algoritma SVR mampu memodelkan data *time series* berdimensi tinggi [9] dan dapat mengatasi overfitting. Beberapa kernel yang populer pada SVR yaitu kernel *linier*, *polinomial*, dan *Radial Basis Function* (RBF) atau *Gaussian* [19]. Namun SVR memiliki kekurangannya, yaitu terdapat banyak parameter yang kompleks sehingga sulit untuk memilih parameter yang tepat untuk mendapatkan hasil pemodelan yang optimal [10].

$$f(x) = w^T \varphi(x) + b \tag{2}$$

2.5.3 Artificial Neural Network

Artificial Neural Network (ANN) merupakan algoritma yang terinspirasi dari struktur jaringan saraf biologis manusia [7]. Algoritma ini memiliki arsitektur yang terdiri dari lapisan masukan (*input*), satu atau lebih lapisan tersembunyi (*hidden*), dan lapisan keluaran (*output*). Algoritma ANN fleksibel dan adaptif [20], karena dapat diimplementasikan pada kasus klasifikasi dan regresi [21]. Tidak hanya itu, ANN bahkan mampu menangani masalah prediksi non-linier yang kompleks [11], [12]. Namun ANN memerlukan data pelatihan yang lebih banyak untuk menangkap pola pada data dengan baik dan memerlukan waktu pelatihan yang lebih lama. Selain itu, banyaknya konfigurasi parameter membuat pengguna sulit untuk menemukan parameter yang tepat, terutama pada jumlah neuron dan lapisannya [7].

$$\hat{y} = \sigma(w_0 + \sum_{i=1}^m x_i w_i) \tag{3}$$

2.6 Evaluasi

2.6.1 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) merupakan suatu parameter nilai akar kuadrat dari *Mean Square Error* (MSE) yang digunakan pada peramalan untuk menghitung nilai kesalahan atau perbedaan antara data prediksi dengan data sebenarnya [22], [23]. Semakin rendah nilai RMSE yang dihasilkan maka hasil prediksi lebih akurat. Untuk menghitung nilai RMSE dapat menggunakan persamaan 3.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{4}$$

2.6.2 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) merupakan rata-rata diferensiasi absolut antara nilai peramalan dan aktual, yang dinyatakan sebagai persentase [22], [23]. Nilai persentase yang rendah menunjukkan hasil prediksi yang semakin baik. Untuk menghitung nilai RMSE dapat menggunakan persamaan 4.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \tag{5}$$

2.7 Analisis Perbandingan

Pada tahap ini, kemampuan prediksi dari setiap model akan dianalisis dan dibandingkan untuk mendapatkan model terbaik. Model terbaik akan dipilih berdasarkan hasil evaluasi dengan nilai RMSE dan MAPE yang paling rendah. Selain itu, hasil prediksi setiap model terhadap data aktual juga akan disajikan melalui visualisasi.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dilakukan melalui observasi dan mengambil data secara langsung dengan rentang waktu 3 tahun terakhir, mulai dari Januari 2021 hingga Desember 2023 dengan total 36 baris data. Data yang digunakan terdiri dari tiga jenis obat yaitu Amoxicilin, Antasida, dan Paracetamol. Data tiga jenis obat yang telah dikumpulkan dapat dilihat pada tabel 1, 2, dan 3.

Tabel 1. Data Obat Amoxicilin

Bulan	Stok	Terjual	Keterangan
Jan-21	500	400	Cukup
Feb-21	500	500	Cukup
Mar-21	500	570	Tidak cukup
Apr-21	580	600	Tidak cukup
...
Des-23	1300	1500	Tidak cukup

Tabel 2. Data Obat Antasida

Bulan	Stok	Terjual	Keterangan
Jan-21	400	420	Tidak cukup
Feb-21	450	430	Cukup
Mar-21	450	460	Tidak cukup
Apr-21	500	480	Cukup
...
Des-23	1115	1200	Tidak cukup

Tabel 3. Data Obat Paracetamol

Bulan	Stok	Terjual	Keterangan
Jan-21	500	550	Tidak cukup
Feb-21	550	600	Tidak cukup
Mar-21	600	620	Tidak cukup
Apr-21	620	650	Tidak cukup
...
Des-23	1020	1045	Tidak cukup

Tabel 1, 2, dan 3 menyajikan sampel data dari tiga jenis obat yang digunakan dalam penelitian ini. Data tersebut menunjukkan tren peningkatan penjualan obat setiap tahunnya, meskipun obat tersebut juga sering mengalami kehabisan stok di apotek.

3.2 Pre-processing

Tahapan awal dari *pre-processing* yaitu melakukan *cleaning* dengan menghapus kolom yang tidak relevan. Selanjutnya ekstraksi fitur dilakukan untuk membentuk rangkaian (*sequence*) deret waktu dengan membentuk fitur baru yang berisikan informasi tentang data stok dan penjualan obat bulan sebelumnya. Pembentukan fitur temporal juga dilakukan untuk menangkap pola yang berkaitan dengan perubahan waktu. Terakhir data dinormalisasi menggunakan teknik *min-max normalization* sehingga nilai pada data akan dalam skala 0 hingga 1. Berikut adalah data yang telah melalui tahap *pre-processing*.

Tabel 4. Hasil Preprocessing Data Obat Amoxicilin

Bulan	Tahun	Stok_Bln_Lalu	Penjualan_Bln_Lalu	Penjualan
0,091	0,000	0,000	0,000	0,000
0,182	0,000	0,117	0,000	0,048
0,273	0,000	0,167	0,133	0,079
0,364	0,000	0,200	0,167	0,103
...
0,727	1,000	1,000	1,000	1,000

Tabel 5. Hasil Preprocessing Data Obat Antasida

Bulan	Tahun	Stok_Bln_Lalu	Penjualan_Bln_Lalu	Penjualan
0,091	0,000	0,000	0,000	0,000
0,182	0,000	0,016	0,077	0,046
0,273	0,000	0,062	0,077	0,076
0,364	0,000	0,093	0,154	0,122
...
0,727	1,000	1,000	1,000	1,000

Tabel 6. Hasil Preprocessing Data Obat Paracetamol

Bulan	Tahun	Stok_Bln_Lalu	Penjualan_Bln_Lalu	Penjualan
0,091	0,000	0,000	0,000	0,000
0,182	0,000	0,111	0,100	0,050
0,273	0,000	0,156	0,200	0,125
0,364	0,000	0,222	0,240	0,125
...
0,727	1,000	1,000	1,000	1,000

Tabel 4, 5, dan 6 menyajikan sampel data tiga jenis obat yang telah dilakukan *pre-processing*. Terlihat bahwa data sudah sesuai dengan format yang membentuk fitur temporal (Bulan dan Tahun), ekstraksi fitur yang membentuk rangkaian deret waktu (Stok_Bln_Lalu dan Penjualan_Bln_Lalu) dan dalam rentang skala yang sama (0 hingga 1).

3.3 Pembagian Data

Teknik *hold-out validation* diimplementasikan untuk membagi dataset dengan persentase 80% sebagai data *training* dan 20% sebagai data *testing*. Hasilnya dari total 36 data, sebanyak 29 data dengan rentang waktu Januari 2021-Mei 2023 dialokasikan sebagai data *training* dan sebanyak 7 data dengan rentang waktu Juni 2023-Desember 2023 dialokasikan sebagai data *testing*.

3.4 Pemodelan

3.4.1 Linear Regression

Pada pemodelan *linear regression*, terdapat 4 variabel independen (X) yang digunakan untuk memprediksi variabel dependen (Y). Keempat variabel independen tersebut adalah Bulan, Tahun, Stock_Bln_Lalu, dan Penjualan_Bln_Lalu, sedangkan variabel dependen (Y) dalam kasus ini adalah Penjualan. Penting untuk dicatat bahwa konsep ini tidak hanya terbatas pada *linear regression*, namun juga berlaku pada beberapa algoritma yang nantinya akan diimplementasikan pada penelitian ini.

3.4.2 Support Vector Regression

Pada algoritma SVR, model akan dilatih dengan beberapa percobaan (*hyperparameter tuning*) pada parameter kernel, C, gamma, dan epsilon untuk menghasilkan model dengan kemampuan terbaik. Informasi *hyperparameter* yang digunakan dilihat pada tabel 7.

Tabel 7. Support Vector Regression Hyperparameter

Parameter	Nilai
Kernel	Linear, Polynomial, RBF
C	0.1, 0.25, 0.5, 1, 2, 3
Gamma	0.1, 0.25, 0.5, 1, 2, 3
Epsilon	0.001, 0.01, 0.1

Tabel 7 menyajikan *hyperparameter* SVR yang akan diuji, dengan aspek utama pada pengujian parameter kernel. Parameter tersebut berperan penting dalam memetakan data non-linear ke dalam ruang fitur berdimensi tinggi, sehingga memungkinkan pemisahan data secara linear. Terdapat 3 jenis kernel yang diujikan, yaitu kernel Linear, Polynomial, dan Radial Basis Function (RBF).

3.4.3 Artificial Neural Network

Arsitektur yang digunakan pada ANN terdiri dari 2 hidden layer dengan 50 neuron di setiap layer-nya. Proses pelatihan pada ANN akan dilakukan menggunakan 4 batch size dengan 200 epoch. Selain itu, model juga akan dilatih dengan beberapa percobaan (*hyperparameter tuning*). Informasi *hyperparameter* yang digunakan dilihat pada tabel 8.

Tabel 8. Artificial Neural Network Hyperparameter

Parameter	Nilai
Optimizer	Adam, SGD, RMSProp
Learning Rate	0.001, 0.0001
Batch Size	4
Epoch	200

Tabel 8 menyajikan *hyperparameter* ANN yang akan diuji, dengan aspek utama pada parameter *optimizer* dan *learning rate*. Parameter tersebut berperan penting dalam memetakan data non-linear ke dalam ruang fitur berdimensi tinggi, sehingga memungkinkan pemisahan data secara linear. Terdapat 3 jenis kernel yang diujikan, yaitu kernel Linear, Polynomial, dan Radial Basis Function (RBF).

3.5 Evaluasi

3.5.1 Linear Regression

Model yang telah dilatih kemudian akan dievaluasi untuk melihat kemampuannya melakukan prediksi. Hasil evaluasi dengan nilai kesalahan (*error*) yang semakin rendah menunjukkan bahwa model yang dikembangkan memberikan hasil yang semakin baik. Hasil evaluasi model LR berdasarkan metrik RMSE dan MAPE pada setiap jenis obat ditampilkan pada Tabel 9.

Tabel 9. Hasil Evaluasi Model Linear Regression

Nama Obat	RMSE	MAPE (%)
Amoxicillin	116.18	4.96
Paracetamol	22.41	1.94
Antasida	28.89	1.67

Tabel 9 menunjukkan hasil evaluasi model LR dengan kategori prediksi sangat baik. Hal tersebut didasari oleh nilai MAPE yang didapatkan pada setiap jenis obat berada dibawah 10% meskipun tanpa melakukan konfigurasi parameter lebih lanjut.

3.5.2 Support Vector Regression

Melalui hasil evaluasi model SVR, didapatkan bahwa kernel Linear lebih dominan memberikan hasil terbaik pada setiap jenis obat, terutama Amoxicilin dan Antasida. Sementara pada jenis obat Paracetamol, Kernel RBF mampu memberikan performa terbaik dibandingkan kernel lainnya. Selain itu, kernel Polynomial menghasilkan performa terburuk hampir pada semua jenis obat.

Tabel 10. Hasil Evaluasi Model Support Vector Regression

Nama Obat	Kernel	Parameter	RMSE	MAPE (%)
Amoxicillin	Linear	C=3; Epsilon=0.001; Gamma=0.1	107.62	4.77
	RBF	C=2; Epsilon=0.001; Gamma=0.1	146.21	6.63
	Polynomial	C=0.5; Epsilon=0.01; Gamma=0.25	91.38	6.69
Antasida	Linear	C=3; Epsilon=0.001; Gamma=0.1	32.08	1.63
	RBF	C=3; Epsilon=0.001; Gamma=0.1	56.78	3.75
	Polynomial	C=0.25; Epsilon=0.01; Gamma=0.25	63.95	4.76
Paracetamol	Linear	C=0.25; Epsilon=0.01; Gamma=0.1	15.03	1.28
	RBF	C=2; Epsilon=0.001; Gamma=0.25	14.32	1.26
	Polynomial	C=0.5; Epsilon=0.1; Gamma=0.25	41.17	3.70

Tabel 10 menunjukkan bahwa model SVR memiliki kategori prediksi sangat baik dengan nilai MAPE yang didapatkan pada setiap jenis obat berada dibawah 10%. Selain itu, hasil pengamatan menunjukkan algoritma SVR lebih dominan memberikan hasil terbaik pada kernel Linear, kemudian diikuti oleh kernel RBF. Sementara itu kernel Polynomial memberikan hasil terburuk hampir pada setiap jenis obat.

3.5.3 Artificial Neural Network

Berdasarkan hasil evaluasi model ANN, terlihat bahwa optimizer RMSprop mampu memberikan hasil yang lebih baik dibandingkan dengan optimizer Adam pada setiap jenis obat. Selain itu, hasil pengamatan menunjukkan semakin kecil nilai learning rate menunjukkan pemodelan yang dihasilkan juga semakin baik. Hal tersebut terlihat dari nilai kesalahan (*error*) yang dihasilkan oleh model yang relatif lebih kecil. Berikut adalah rincian dari hasil evaluasi model ANN.

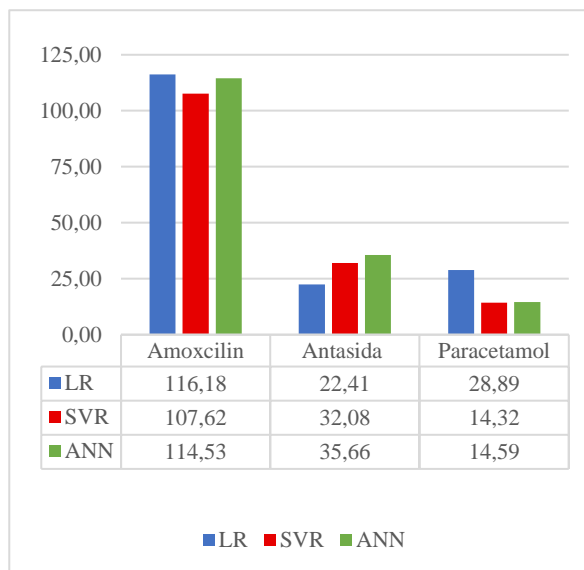
Tabel 11. Hasil Evaluasi Model Artificial Neural Network

Nama Obat	Optimizer	Learning Rate	RMSE	MAPE
Amoxicillin	Adam	0.001	135.77	5.92
		0.0001	116.56	5.15
		RMSprop	0.001	128.70
Antasida	Adam	0.0001	114.53	4.73
		0.001	37.60	1.65
		0.0001	36.10	1.63
Paracetamol	RMSprop	0.001	42.80	2.30
		0.0001	35.66	1.56
		0.001	21.78	1.82
Paracetamol	Adam	0.001	21.78	1.82
		0.0001	16.42	1.35
		RMSprop	0.001	20.47
		0.0001	14.59	1.17

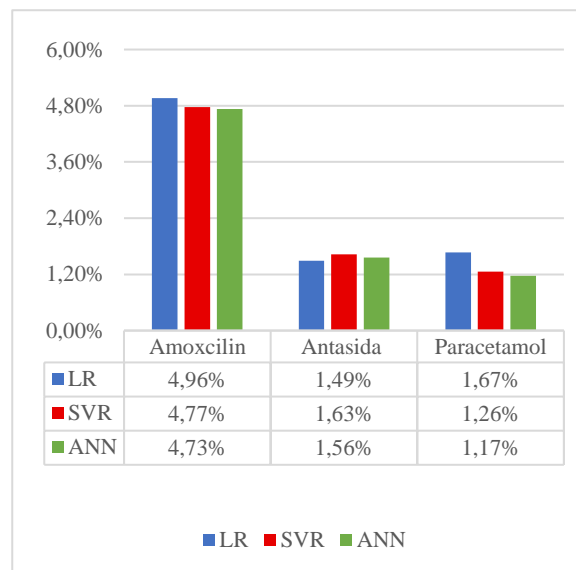
Tabel 11 menunjukkan bahwa model ANN memiliki kategori prediksi sangat baik dengan nilai MAPE yang didapatkan pada setiap jenis obat berada dibawah 10%. Hasil pengamatan menunjukkan performa ANN memberikan hasil yang lebih baik jika menggunakan optimizer RMSprop dibandingkan Adam. Selain itu, nilai learning rate yang semakin kecil mampu memberikan hasil yang lebih maksimal.

3.6 Analisis Perbandingan

Setelah melakukan evaluasi, tahapan selanjutnya adalah melakukan perbandingan algoritma untuk mendapatkan model terbaik. Setiap model akan dibandingkan berdasarkan tingkat kesalahan (*error*) terendah berdasarkan metrik RMSE dan MAPE yang telah didapatkan sebelumnya. Hasil perbandingan dari tiga model algoritma pada penelitian ini dapat dilihat pada Gambar 2 dan Gambar 3.



Gambar 2. Perbandingan Berdasarkan Nilai RMSE

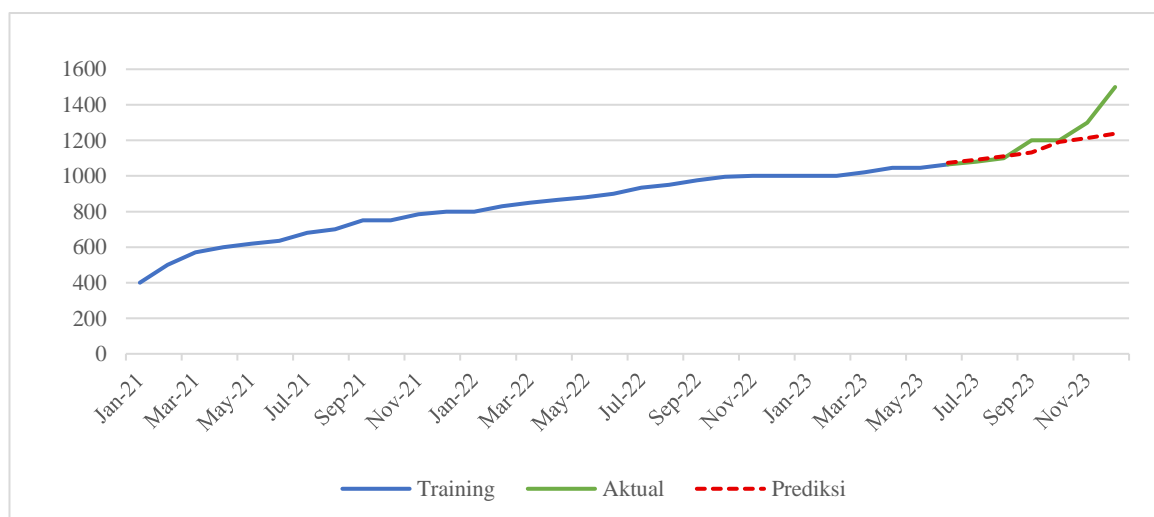


Gambar 3. Perbandingan Berdasarkan Nilai MAPE

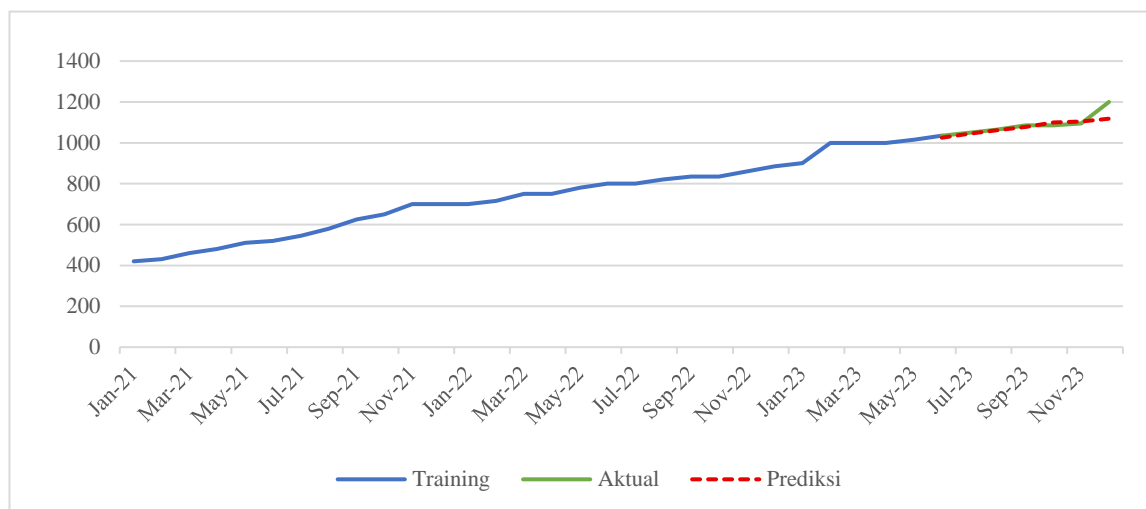
Gambar 2 dan 3 menyajikan perbandingan nilai RMSE dan MAPE yang menghasilkan beberapa informasi. Jika hasil perbandingan berfokus pada metrik RMSE, maka model SVR sedikit lebih unggul dibandingkan dengan model ANN. Pada jenis obat Amoxicilin, SVR mendapatkan nilai RMSE sebesar 107.62 sedangkan ANN sebesar 114.53. Kemudian pada jenis obat Antasida, SVR memiliki nilai RMSE sebesar 32.08 sedangkan ANN sebesar 35.66. Begitu juga pada jenis obat Paracetamol, dimana SVR memiliki nilai RMSE sebesar 14.32 sedangkan ANN sebesar 14.59. Namun hal sebaliknya terjadi jika hasil perbandingan berfokus pada metrik MAPE, dimana ANN sedikit lebih unggul dibandingkan dengan SVR. Pada jenis obat Amoxicilin, ANN memiliki nilai MAPE sebesar 4.73% sedangkan SVR sebesar 4.77%. Kemudian pada jenis obat Antasida, ANN mendapatkan nilai MAPE sebesar 1.56% sedangkan SVR sebesar 1.63%. Ini juga berlaku pada jenis obat Paracetamol, dimana ANN memiliki nilai MAPE sebesar 1.17% sedangkan SVR sebesar 1.26%.

Beberapa pertimbangan dapat dilakukan untuk menghadapi kasus tersebut berdasarkan konteks dan kebutuhan analisis. Jika kebutuhan analisis menekankan pada akurasi model secara keseluruhan, maka RMSE dapat menjadi pilihan yang lebih baik. Namun jika kebutuhan analisis menekankan pada akurasi model secara relatif dan juga membandingkan hasil pada dataset yang berbeda, maka MAPE dapat menjadi pilihan yang lebih baik. Untuk itu pada penelitian ini fokus utama analisis perbandingan yaitu pada metrik RMSE. Hal ini dikarenakan pada penelitian ini dataset yang digunakan masih sama dan juga lebih menekankan akurasi model secara keseluruhan. Oleh karena itu, berdasarkan pertimbangan tersebut maka SVR merupakan model terbaik mengungguli ANN dan juga LR untuk kasus prediksi tiga jenis data obat.

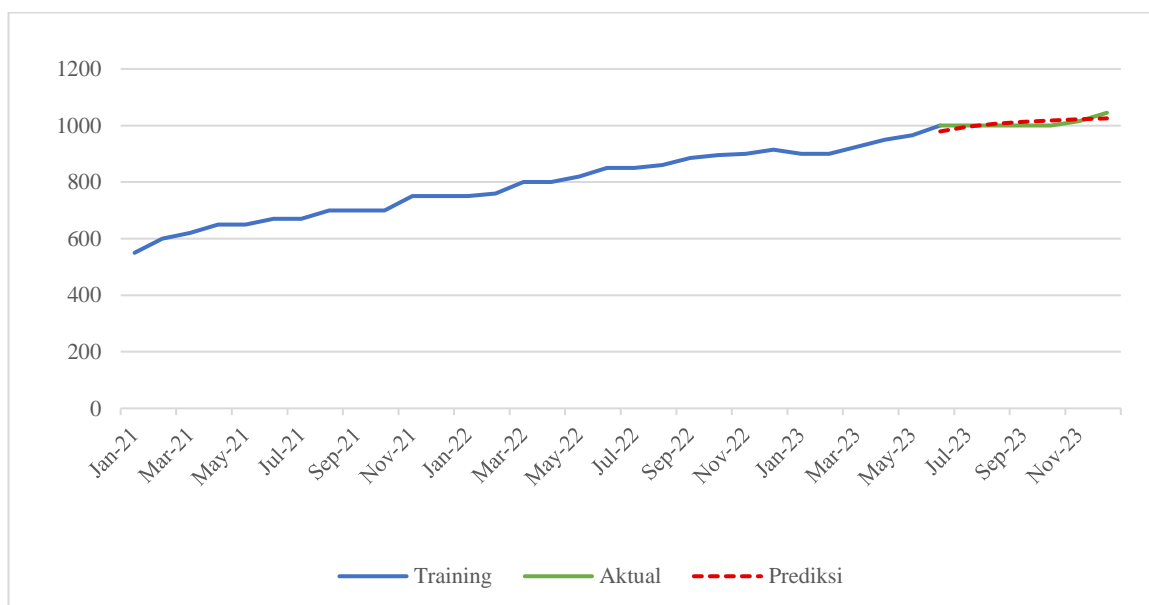
Setelah mendapatkan model terbaik, tahapan selanjutnya adalah melakukan visualisasi hasil prediksi terhadap data aktual untuk melihat seberapa baik hasil prediksi menangkap pola dan trend dari data. Ketika hasil prediksi semakin berdekatan dan mengikuti pola dari data aktual, maka hasil prediksi dapat dikatakan sangat baik. Hasil visualisasi prediksi pada ketiga data penjualan obat dapat dilihat pada gambar dibawah.



Gambar 4. Hasil Prediksi Penjualan Obat Amoxiciline Terhadap Data Aktual



Gambar 5. Hasil Prediksi Penjualan Obat Antasida Terhadap Data Aktual



Gambar 6. Hasil Prediksi Penjualan Obat Paracetamol Terhadap Data Aktual

Berdasarkan gambar 4, 5 dan 6, terlihat bahwa hasil prediksi dari model mampu mengikuti pola dan tren pada data aktual dengan baik, terutama pada data jenis obat Antasida dan Paracetamol. Sementara pada jenis obat Amoxicilin, hasil prediksi model sedikit meleset pada bulan Desember 2023, dimana data aktual pada bulan tersebut terjadi peningkatan yang cukup signifikan. Meskipun begitu secara keseluruhan hasil prediksi model pada setiap jenis obat sudah sangat baik. Namun penambahan jumlah data dapat dipertimbangkan untuk memberikan pemahaman yang lebih pada model untuk mempelajari pola dan tren pada data aktual sehingga dapat meningkatkan kemampuan model dan akurasi prediksi.

4. KESIMPULAN

Kemampuan tiga algoritma regresi yang diimplementasikan yaitu Linear Regression (LR), Support Vector Regression (SVR) dan Artificial Neural Network (ANN) memberikan hasil yang beragam pada tiga jenis obat yang diujikan. Pada implementasinya, algoritma LR lebih sederhana dan mudah karena tidak memerlukan konfigurasi parameter yang banyak seperti SVR dan ANN. Hasil pengamatan menunjukkan algoritma SVR lebih dominan memberikan hasil terbaik pada kernel Linear, kemudian diikuti oleh kernel RBF. Sementara itu kernel Polynomial memberikan hasil terburuk hampir pada setiap jenis obat. Selanjutnya, performa ANN lebih baik jika menggunakan optimizer RMSprop dibandingkan Adam dan juga learning rate yang semakin kecil memberikan hasil yang lebih maksimal. Melalui hasil analisis dan perbandingan yang dilakukan, algoritma SVR memiliki kemampuan yang lebih baik dibandingkan ANN dan LR jika konteks dan analisis berfokus pada metrik RMSE. Sementara itu, jika konteks dan analisis berfokus pada metrik MAPE, maka ANN lebih unggul dibandingkan algoritma lainnya. Meskipun LR tidak lebih unggul dalam hal



metrik apapun, nyatanya algoritma tersebut termasuk SVR, dan ANN memiliki nilai MAPE dibawah 10%, dimana hasil prediksinya masih dalam kategori sangat akurat. Selain itu, melalui visualisasi yang dilakukan menunjukkan hasil prediksi dari seluruh model yang diusulkan mampu mengikuti pola dan tren pada data aktual dengan baik. Namun penambahan jumlah data dapat dipertimbangkan untuk memberikan pemahaman yang lebih pada model untuk mempelajari pola dan tren pada data aktual sehingga dapat meningkatkan kemampuan model dan akurasi prediksi.

REFERENCES

- [1] A. Prasetyo, S. Salahuddin, and A. Amirullah, "Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda," *Jurnal Infomedia*, vol. 6, no. 2, p. 76, 2021, doi: 10.30811/jim.v6i2.2343.
- [2] A. D. Sidik and A. Ansawarman, "Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning," *Formosa Journal of Multidisciplinary Research*, vol. 1, no. 3, pp. 559–568, 2022, doi: 10.55927/fjmr.v1i3.745.
- [3] R. Ternando, J. Wahyudi, and D. Suranti, "Application of the Multiple Linear Regression Method in Forecasting the Amount of Drug Supply at the Health Center," *Jurnal Komputer, Informasi dan Teknologi (JKOMITEK)*, vol. 2, no. 2, pp. 621–628, 2022, doi: 10.53697/jkomitek.v2i2.1043.
- [4] A. M. A. Rusdy, P. Purnawansyah, and H. Herman, "Penerapan Metode Regresi Linear Pada Prediksi Penawaran dan Permintaan Obat Studi Kasus Aplikasi Point Of Sales," *Buletin Sistem Informasi dan Teknologi Islam*, vol. 3, no. 2, pp. 121–126, 2022, doi: 10.33096/busiti.v3i2.1130.
- [5] P. Bholra and S. Bhardwaj, "Estimation of solar radiation using support vector regression," *Journal of Information and Optimization Sciences*, vol. 40, no. 2, pp. 339–350, 2019, doi: 10.1080/02522667.2019.1578093.
- [6] Q. Quan, Z. Hao, H. Xifeng, and L. Jingchun, "Research on water temperature prediction based on improved support vector regression," *Neural Computing and Applications*, vol. 34, no. 11, pp. 8501–8510, 2022, doi: 10.1007/s00521-020-04836-4.
- [7] H. Putra and N. U. Walmi, "Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 2, pp. 100–107, 2020.
- [8] M. Masrurroh and K. F. Mauladi, "Penerapan Metode Regresi Linear Berganda Dalam Sistem Prediksi Nilai Ujian Nasional Siswa Smp," *Jurnal Teknika*, vol. 12, no. 1, p. 1, 2020, doi: 10.30736/jt.v12i1.393.
- [9] S. Saadah, F. Z. Zai, and H. H. Z, "Support Vector Regression (SVR) Dalam Memprediksi Harga Minyak Kelapa Sawit di Indonesia dan Nilai Tukar Mata Uang EUR/USD," *Journal of Computer Science and Informatics Engineering (J-Cosine)*, vol. 5, no. 1, pp. 85–92, 2021, doi: 10.29303/jcosine.v5i1.403.
- [10] G. H. Saputra, A. H. Wigena, and B. Sartono, "Penggunaan Support Vector Regression Dalam Pemodelan Indeks Saham Syariah Indonesia Dengan Algoritme Grid Search," *Indonesian Journal of Statistics and Its Applications*, vol. 3, no. 2, pp. 148–160, 2019, doi: 10.29244/ijsa.v3i2.172.
- [11] N. Tampati, "Prediksi Stok Obat pada Apotik Total Life Clinic Menggunakan Model Kombinasi Artificial Neural Network dan ARIMA," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 1, pp. 49–58, 2022, doi: 10.35957/jatisi.v9i1.1373.
- [12] Z. Jannah and K. Budayawan, "Pengembangan Sistem Cerdas Berbasis Data Mining untuk Meningkatkan Akurasi Prediksi Kebutuhan Obat di Puskesmas Parit Rantang," *Al-DYAS*, vol. 3, no. 1, pp. 467–479, 2024.
- [13] S. Agnesti, A. Nazir, I. Iskandar, E. Budianita, and I. Afrianty, "Perbandingan Algoritma Triple Exponential Smoothing dan Support Vector Regression dalam Prediksi Pemakaian Obat di Puskesmas," *SISTEMASI: Jurnal Sistem Informasi*, vol. 12, no. 3, pp. 996–1006, 2023, doi: 10.32520/stmsi.v12i3.3499.
- [14] S. Pratista, A. Nazir, I. Iskandar, E. Budianita, and I. Afrianty, "Perbandingan Prediksi Obat Berdasarkan Pemakaian Menggunakan Algoritma Single Moving Average dan Support Vector Regression," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, pp. 1860–1868, 2023, doi: 10.30865/mib.v7i4.6859.
- [15] R. Novita, I. Yani, and G. Ali, "Sistem Prediksi untuk Penentuan Jumlah Pemesanan Obat Menggunakan Regresi Linier," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 1, pp. 62–70, 2022, doi: 10.57152/malcom.v2i1.198.
- [16] N. E. Warestika, D. Sugiarto, and T. Siswanto, "Business Intelligence Design for Data Visualization and Drug Stock Forecasting," *Intelmatix*, vol. 1, no. 1, pp. 9–15, 2021, doi: 10.25105/itm.v1i1.7407.
- [17] A. Rahmadeyan, Mustakim, M. Erkamim, I. Ahmad, Sepriano, and S. Aziz, "Air Pollution Prediction Using Long Short-Term Memory Variants," in *Advances in Intelligent Computing Techniques and Applications*, Springer, 2024, pp. 122–132.
- [18] R. E. Cahyono and J. P. Sugiono, "Analisis Kinerja Metode Support Vector Regression (SVR) dalam Memprediksi Indeks Harga Konsumen," *Jurnal Teknologi Informasi dan Multimedia*, vol. 1, no. 2, pp. 106–116, 2019, doi: 10.35746/jtim.v1i2.22.
- [19] Y. Yang, J. Che, C. Deng, and L. Li, "Sequential grid approach based support vector regression for short-term electric load forecasting," *Applied Energy*, vol. 238, pp. 1010–1021, 2019, doi: 10.1016/j.apenergy.2019.01.127.
- [20] N. Sepriyanti, R. S. Nahampun, M. H. Zikri, I. Ambarani, and A. Rahmadeyan, "Implementation of K-Means Clustering to Group Poverty Levels in Riau Province," in *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 2022, pp. 59–65.
- [21] A. Rahmadeyan, Mustakim, I. Ahmad, A. D. Alexander, and A. Rahman, "Phishing Website Detection with Ensemble Learning Approach Using Artificial Neural Network and AdaBoost," in *2023 International Conference on Information Technology Research and Innovation (ICITRI)*, 2023, pp. 162–166, doi: 10.1109/ICITRI59340.2023.10249799.
- [22] A. Rahmadeyan and Mustakim, "Long Short-Term Memory and Gated Recurrent Unit for Stock Price Prediction," *Procedia Computer Science*, vol. 234, pp. 204–212, 2024, doi: 10.1016/j.procs.2024.02.167.
- [23] N. T. Luchia, E. Tasia, I. Ramadhani, A. Rahmadeyan, and R. Zahra, "Performance Comparison Between Artificial Neural Network, Recurrent Neural Network and Long Short-Term Memory for Prediction of Extreme Climate Change," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 1, no. 2, pp. 62–70, 2024, doi: 10.57152/predatecs.v1i2.864.