

Prediksi Curah Hujan di Jakarta Menggunakan Model Hybrid (DWT-SVR-Prophet)

Mutiara Ramadita¹, Mahmudi², Madona Yunita Wijaya³

mutiararmadita@gmail.com¹, mahmudi@uinjkt.ac.id², madona@uinjkt.ac.id³

^{1,2,3} Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia

Informasi Artikel

Diterima : 4 Agu 2024
Direview : 29 Agu 2024
Disetujui : 3 Sep 2024

Kata Kunci

DWT, Hybrid, Prediksi
Curah Hujan, Prophet,
SVR

Abstrak

Curah hujan memiliki pengaruh signifikan dalam perencanaan dan pengelolaan berbagai industri seperti kehutanan, pertanian, dan sumber daya air. Penelitian ini bertujuan untuk mengembangkan model Hybrid yang menggabungkan model *Discrete Wavelet Transform* (DWT), *Support Vector Regression* (SVR) dan *Prophet* untuk memprediksi curah hujan di wilayah Jakarta dengan lebih akurat. Dengan menggunakan DWT, data curah hujan dibagi menjadi komponen frekuensi tinggi dan rendah, kemudian hasil prediksi dari masing-masing model digabungkan. Di Stasiun Kemayoran model Hybrid (SVR Frekuensi Tinggi dan Prophet Frekuensi Rendah) memberikan kinerja terbaik dengan SMAPE: 36.44%. Di Stasiun Tanjung Priok, model non-Hybrid memberikan hasil terbaik, dengan Prophet tanpa DWT mencapai SMAPE: 29.82%. Penelitian ini memberikan pemahaman yang jelas tentang seberapa efektif model prediksi curah hujan di Jakarta, yang membantu perencanaan dan pengelolaan sumber daya air.

Keywords

DWT, Hybrid, Prophet,
Rainfall Prediction, SVR

Abstract

Rainfall has a significant influence in the planning and management of various industries such as forestry, agriculture, and water resources. This research aims to develop a Hybrid model that combines Discrete Wavelet Transform (DWT), Support Vector Regression (SVR) and Prophet models to predict rainfall in the Jakarta area more accurately. Using DWT, rainfall data is divided into high and low frequency components, then the prediction results from each model are combined. At Kemayoran Station, the Hybrid model (High Frequency SVR and Low Frequency Prophet) provides the best performance with SMAPE: 36.44%. At Tanjung Priok Station, the non-Hybrid model gave the best results, with Prophet without DWT achieving SMAPE: 29,82%. This study provides a clear understanding of how effective rainfall prediction models are in Jakarta, which helps water resources planning and management.

A. Pendahuluan

Curah hujan merupakan komponen yang sangat penting dari siklus hidrologi global yang sangat mempengaruhi kehidupan manusia dan lingkungan [1]. Curah hujan didefinisikan sebagai jumlah air yang turun ke permukaan bumi dalam bentuk hujan, salju, atau es yang berperan penting dalam keseimbangan air, pembentukan iklim regional dan ketersediaan sumber daya air [2]. Di wilayah tropis seperti Indonesia, khususnya Jakarta, variasi curah hujan memiliki dampak luas, mulai dari ketahanan pangan hingga pengelolaan risiko bencana [3]. Pola curah hujan di Jakarta dipengaruhi oleh berbagai faktor seperti musim monsun, El Niño Southern Oscillation (ENSO), dan Indian Ocean Dipole (IOD), yang menyebabkan fluktuasi signifikan dalam distribusi curah hujan baik secara temporal maupun spasial [4], [5]. Oleh karena itu, mengembangkan metode prediksi curah hujan yang lebih canggih dan akurat menjadi prioritas untuk mengurangi risiko bencana dan mengelola sumber daya air secara berkelanjutan di daerah perkotaan seperti Jakarta [5].

Penelitian sebelumnya telah menunjukkan berbagai metode untuk memprediksi curah hujan di Jakarta, dengan mempertimbangkan variabel eksogen. Penelitian yang menggabungkan teknik ensemble averaging dengan ARIMA dan ARIMAX menunjukkan bahwa kelembaban adalah variabel eksogen yang signifikan dalam meramalkan curah hujan [6]. Walaupun tidak memberikan peningkatan yang besar dalam akurasi dibanding model ARIMA atau ARIMAX terbaik, penggunaan ensemble averaging membantu mengatasi masalah dalam memilih model. Di Stasiun Kemayoran, model ARIMAX (0,0,1) memiliki nilai SMAPE sebesar 27.59%, yang lebih rendah dari Stasiun Tanjung Priok dengan nilai SMAPE sebesar 37.83%. Sementara itu, gabungan ARIMAX-ANN memberikan hasil yang lebih akurat daripada model ARIMAX konvensional. Sementara itu, penelitian yang menggunakan model Hybrid ARIMAX-ANN menunjukkan bahwa variabel eksogen yang signifikan dalam pemodelan curah hujan menggunakan model ARIMAX adalah kelembaban, sedangkan suhu tidak diikutsertakan karena tidak memiliki pengaruh signifikan [7]. Model Hybrid ARIMAX-ANN terbukti lebih akurat dibandingkan model ARIMAX dalam memodelkan curah hujan di Jakarta. Di Stasiun Kemayoran, model dengan satu hidden layer memberikan hasil terbaik, sedangkan di Stasiun Tanjung Priok, dua hidden layer menghasilkan akurasi yang lebih baik. Model Hybrid ARIMAX-ANN di Stasiun Kemayoran menghasilkan MAPE sebesar 21.145%, yang lebih baik dibandingkan di Stasiun Tanjung Priok dengan MAPE sebesar 37.416%.

Meskipun kedua penelitian tersebut menunjukkan hasil yang baik dengan menambahkan variabel eksogen seperti kelembaban, penelitian ini memilih fokus hanya pada data curah hujan. Dengan menggunakan data curah hujan, penelitian ini bertujuan untuk lebih memahami pola curah hujan itu sendiri. Ini didasarkan bahwa data curah hujan sudah memiliki informasi yang cukup untuk membuat prediksi yang akurat, tanpa perlu melibatkan variabel eksogen.

Penelitian terus dilakukan untuk meningkatkan akurasi prediksi curah hujan. Salah satu metodenya adalah dengan menggunakan *Discrete Wavelet Transform* (DWT) yang menggunakan fungsi wavelet untuk transformasi data. Dengan DWT data curah hujan dibagi menjadi dua bagian: koefisien detail (frekuensi tinggi) dan koefisien aproksimasi (frekuensi rendah), ini membantu menemukan pola yang

tidak terlihat dalam data asli. Haar Wavelet, salah satu jenis wavelet yang digunakan Husain dan Amran [8], terbukti efektif untuk menganalisis dan memprediksi data deret waktu seperti kelembaban udara dan curah hujan. Disisi lain, *Support Vector Regression* (SVR) juga menunjukkan potensi yang besar dalam prediksi curah hujan. *Support Vector Regression* (SVR) adalah perluasan dari *Support Vector Machine* (SVM) dalam menangani kasus regresi dengan output bilangan riil atau kontinyu. Metode ini efektif dalam menggambarkan hubungan antara variabel linear dan non-linear, serta memiliki kemampuan untuk mengatasi overfitting. Khairunisa dkk menunjukkan bahwa prediksi menggunakan SVR menunjukkan reliabilitas dan stabilitas hyperplane yang tinggi. Selain itu, penelitian oleh Asyraf [9] juga membuktikan keakuratan SVR dalam meramalkan curah hujan di Kota Medan.

Metode *Prophet* adalah metode yang dibuat oleh Facebook untuk meramalkan data *time series* dengan mempertimbangkan tren, musiman dan hari libur. Penelitian Sulasikin [10] menjelaskan bahwa model yang terbaik untuk memprediksi curah hujan bulanan di Jakarta Pusat karena memiliki performa yang terbaik dengan *Mean Squared Error* (MSE) dan *Root Mean Squared Error* (RMSE) yang terendah dibandingkan dengan model SARIMA dan LSTM.

Dibutuhkan metode lebih lanjut untuk meningkatkan akurasi prediksi curah hujan, meskipun metode-metode tersebut menunjukkan hasil yang baik ketika digunakan secara individual. Dongsheng [11] menemukan bahwa model Hybrid yang menggabungkan DWT, SVR, dan Prophet dapat meningkatkan akurasi prediksi curah hujan secara signifikan. Tidak seperti model individual, model Hybrid ini efektif dalam menangkap pola-pola kompleks dalam data curah hujan. Tujuan penelitian ini adalah untuk mengembangkan model prediksi curah hujan yang lebih akurat di Jakarta berdasarkan temuan ini. Dengan menggabungkan DWT, SVR, dan Prophet, diharapkan model ini dapat menangkap kompleksitas pola curah hujan di Jakarta dengan lebih baik. Metode ini berpotensi untuk menangkap pola yang lebih kompleks karena menggabungkan analisis frekuensi (DWT), kemampuan menangani non-linearitas (SVR), dan analisis tren serta musiman (Prophet). Hasil dari penelitian ini diharapkan memberikan kontribusi yang berarti dalam pengelolaan sumber daya air dan mitigasi bencana di Jakarta.

B. Metode Penelitian

1. Data

Penelitian ini menggunakan jenis penelitian kuantitatif. Data sekunder yang digunakan dalam penelitian ini adalah data curah hujan bulanan dari Stasiun Meteorologi Kemayoran dan Stasiun Meteorologi Maritim Tanjung Priok dari bulan Januari 2018 hingga bulan Juni 2023. Untuk mengevaluasi kinerja model dan memastikan kemampuan generalisasinya pada data deret waktu, diterapkan metode cross validasi yang dikenal sebagai 'time series cross-validation' atau 'rolling window cross-validation', sejalan dengan pendekatan yang diuraikan oleh Hyndman dan Athanasopoulos [12] dalam buku mereka tentang prinsip dan praktik peramalan. Data dipisahkan menjadi 11 bagian yang berbeda menggunakan metode $k=11$. Validasi dilakukan dengan metode expanding window dimana setiap kali iterasi menambahkan satu fold ke data training dan menggunakan fold berikutnya sebagai data testing. Langkah awal dari cross validasi adalah memperhitungkan ukuran window berdasarkan jumlah data total dan jumlah fold,

dan kemudian dilakukan iterasi dari 1 hingga k-1. Proses cross validation dilakukan sebagai berikut:

- 1) Data training mencakup data dari awal dataset hingga akhir window.
- 2) Data testing terdiri dari data setelah akhir window training hingga akhir window berikutnya.
- 3) Model dilatih menggunakan data training.
- 4) Data testing digunakan untuk memprediksi.
- 5) SMAPE dihitung untuk mengukur akurasi prediksi.

SMAPE digunakan sebagai metrik evaluasi karena mampu menangani nilai nol dan ekstrem serta memberikan persentase error yang simetris, sehingga menghasilkan estimasi kinerja model yang lebih stabil dan dapat diandalkan. Hasil SMAPE dari setiap iterasi akan dihitung rata-ratanya. Metode validasi time series cross-validation ini dapat mengurangi risiko overfitting, memberikan estimasi kinerja model yang lebih akurat, dan menggunakan sifat data curah hujan yang berkaitan dengan waktu saat evaluasi dilakukan.

2. Discrete Wavelet Transform

Transformasi wavelet pertama kali diperkenalkan pada tahun 1980-an oleh Morlet dan Grossman sebagai suatu fungsi matematis yang digunakan untuk merepresentasikan data atau fungsi sebagai alternatif transformasi matematika yang telah ada sebelumnya untuk menangani masalah resolusi. Transformasi wavelet ini memiliki dua jenis dalam pengembangannya, yaitu *continuous Wavelet Transform* (CWT) dan *Discrete Wavelet Transform* (DWT). DWT melakukan dekomposisi data input ke dalam komponen frekuensi yang berbeda tanpa kehilangan informasi penting. DWT merupakan fungsi skala yang mengkonversi data input menjadi fungsi dasar transformasi wavelet (*father* dan *mother wavelet*) [13].

DWT mengolah karakteristik data input dengan menggunakan dua set fungsi, yaitu filter high-pass dan filter low-pass. Keberhasilan proses karakteristik data dapat dipengaruhi dalam pemilihan fungsi dasar transformasi wavelet. DWT dipilih untuk melakukan dekomposisi deret waktu curah hujan karena metodenya mengandalkan proses pemrosesan yang sederhana dan lebih cocok untuk mengolah data *time series* [14]. Persamaan transformasi wavelet adalah sebagai berikut:

$$W(p, q) = 2^{-\left(\frac{p}{2}\right)} \sum_{t=0}^T \psi\left(\frac{t - q \cdot 2^p}{2^p}\right) \quad (1)$$

Dimana ψ adalah fungsi dasar wavelet, t adalah parameter waktu, T adalah panjang sinyal, p adalah parameter skala, q adalah translasi waktu dan $x(t)$ adalah data target.

3. Support Vector Regression

Support Vector Regression (SVR) merupakan pendekatan yang diperluas berdasarkan konsep *Support Vector Machine* (SVM). Pendekatan ini memberikan beberapa keuntungan, termasuk dalam hal minimasi risiko struktural dan

kemampuannya dalam menggunakan ukuran sampel yang kecil. SVR sebenarnya adalah penerapan SVM di bidang regresi [15].

Proses regresinya dapat dijelaskan sebagai berikut:

$$f(x) = w^T \cdot x + b \quad (2)$$

Misalkan terdapat sejumlah data $\{(x_i, y_i), i = 1, 2, \dots, N\}$ dimana $x_i \in R^n$ adalah variabel input dan $y_i \in R$ adalah variabel output. Mengambil pertimbangan sebuah pemetaan sedemikian rupa sehingga $\varphi(x)$ berubah menjadi vektor eigen dari x setelah dipetakan ke dimensi yang lebih tinggi, menghasilkan fungsi regresi linear sebagai berikut:

$$f(x) = w^T \cdot \varphi(x) + b \quad (3)$$

Dimana w adalah koefisien, $b \in R$ adalah deviasi dan $\varphi(x)$ adalah fungsi pemetaan non linear dari ruang berdimensi rendah ke ruang berdimensi tinggi, model dibangun setelah w dan b diperoleh melalui proses pembelajaran.

Berdasarkan kriteria minimasi risiko struktural, masalah tersebut diubah menjadi suatu permasalahan minimasi fungsi objektif R , yang dirumuskan sebagai berikut:

$$\begin{aligned} \min_{(w, e)} R(w, e) &= \frac{1}{2} w^T \cdot w + C \sum_{i=1}^n e_i^2 \\ s \cdot t \cdot y_i &= w^T \varphi(x_i) + b + e_i \end{aligned} \quad (4)$$

dimana $R(w, e)$ adalah fungsi objektif, persamaan $s \cdot t$ menunjukkan batasan yang harus dipenuhi oleh R ketika di minimalkan, $y_i \in R$ adalah variabel output, $e_i \in R$ adalah variabel error yang akan di tentukan melalui model training dan C adalah koefisien penalty yang nilainya lebih besar dari 0.

Untuk menyelesaikan permasalahan minimasi fungsi objektif R di atas, maka dibentuklah fungsi Lagrange L .

$$\begin{aligned} L(w, b, e_i, \alpha_i, \tilde{\alpha}_i) &= R(w, e) \\ &- \sum_{i=1}^n \alpha_i [w^T \varphi(x_i) + b + e_i - y_i] \\ &+ \sum_{i=1}^n \tilde{\alpha}_i [w^T \varphi(x_i) + b + e_i - y_i] \end{aligned} \quad (5)$$

dimana $\tilde{\alpha}_i, \alpha_i$ adalah operator Lagrange multiplier dan definisi w, b, e_i dan y_i sama seperti yang telah dijelaskan sebelumnya. Persamaan tersebut kemudian di

substitusikan ke dalam persamaan (2) dan menghasilkan persamaan sebagai berikut:

$$w = \sum_{i=1}^n (\tilde{\alpha}_i - \alpha_i) \cdot x_i \quad (6)$$

dimana $\tilde{\alpha}_i, \alpha_i$ memiliki karakteristik yang sama seperti yang telah dijelaskan sebelumnya dan dapat membuat sampel dimana $\tilde{\alpha}_i, \alpha_i \neq 0$ yang merupakan support vektor dari SVR.

Dengan memenuhi kondisi Karush-Kuhn-Tucker (KKT) dalam permasalahan dual SVR dan mensubstitusikan rumus yang dihasilkan ke dalam persamaan (2), solusi SVR dapat dihasilkan seperti berikut:

$$f(x) = \sum_{i=1}^n (\tilde{\alpha}_i - \alpha_i) \cdot x_i^T x + b \quad (7)$$

dimana bias dinyatakan dalam $b = y_i - \sum_{i=1}^n (\tilde{\alpha}_i - \alpha_i) \cdot x_i^T x$. Sehingga diperoleh solusi SVR adalah sebagai berikut:

$$f(x) = \sum_{i=1}^n (\tilde{\alpha}_i - \alpha_i) \cdot k(x_i, x) + b \quad (8)$$

4. Prophet

Model Prophet adalah sebuah kerangka kerja untuk prediksi data deret waktu yang dikembangkan oleh Meta (dahulu dikenal sebagai Facebook) pada tahun 2017. Perangkat lunak ini bersifat open source dan dirancang oleh Tim Data Sains Facebook pada tahun 2017 [16], bertujuan untuk menangani karakteristik umum dari data time series. Prophet tersedia sebagai library di Python dan R [17] dan menggunakan model regresi modular sederhana yang biasanya memberikan hasil yang baik dengan pengaturan default.

Prophet menggunakan model time series yang dapat didekomposisi dengan tiga komponen utama yaitu tren (trend), musiman (seasonality), dan liburan (holidays) yang dikombinasikan dengan persamaan berikut [17]:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (9)$$

Dari persamaan diatas $g(t)$ merupakan fungsi tren yang memodelkan perubahan non-periodik dari data time series, $s(t)$ merepresentasikan perubahan periodic (seperti musiman mingguan dan tahunan), $h(t)$ merepresentasikan efek dari hari libur yang terjadi pada jadwal yang berpotensi tidak teratur selama satu hari atau lebih, dan ε_t yaitu istilah kesalahan (error term) yang merepresentasikan setiap perubahan yang tidak diakomodasi oleh model.

1. Tren (Trend)

Prophet menggunakan dua model tren yaitu: *saturating growth model* dan *piecewise linear model*.

1) Saturating Growth Model (non-linear)

Untuk peramalan pertumbuhan, komponen utama dalam proses pembentukan data adalah model yang menggambarkan bagaimana populasi telah tumbuh dan diperkirakan akan terus tumbuh. Berikut adalah persamaan untuk model *piecewise logistic growth*:

$$g(t) = \frac{C(t)}{1 + \exp\left(-(k + a(t)^T \delta)(t - (m + a(t)^T \gamma))\right)} \quad (10)$$

2) Piecewise Linear Model (Linear)

Untuk masalah peramalan yang tidak menunjukkan pertumbuhan yang jenuh (*saturating growth*), maka model trennya adalah:

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \quad (11)$$

dimana k adalah growth rate, δ memiliki rate adjustments, m adalah parameter offset, dan γ_j diatur ke $-s_j \delta_j$ untuk membuat fungsi kontinu.

2. Musiman (Seasonal)

Data deret waktu sering kali menunjukkan pola musiman dengan beberapa periode. Model Prophet mengatasi hal ini menggunakan deret Fourier untuk memberikan model efek periodik yang fleksibel. Berikut adalah persamaannya.

$$s(t) = \sum_{n=1}^N \left[a_n \cos\left(\frac{2\pi n t}{p}\right) + b_n \sin\left(\frac{2\pi n t}{p}\right) \right] \quad (12)$$

dengan parameter $\beta = [a_1, b_1, \dots, a_N, b_N]^T$. Hal ini dilakukan dengan membangun matriks vector musiman untuk setiap nilai t dalam data historis dan masa depan. Sebagai contoh untuk yearly seasonality dengan $P = 365.25$ dan $N = 10$.

$$X(t) = \left[\sin\left(\frac{2\pi(1)t}{365.25}\right), \dots, \cos\left(\frac{2\pi(10)t}{365.25}\right) \right] \quad (13)$$

Sehingga komponen musiman menjadi seperti berikut.

$$s(t) = X(t)\beta \quad (14)$$

Dalam model generative ini ditetapkan $\beta \sim \text{Normal}(0, \sigma^2)$, untuk pemulusan sebelumnya pada musiman. Kemudian P adalah periode reguler dari deret waktu, dan N untuk menerapkan low pass filter pada musiman. Meningkatkan nilai N memungkinkan penyesuaian pola musiman yang lebih cepat, meskipun akan meningkatkan risiko *overfitting*.

3. Hari libur (Holidays & Events)

Dampak hari libur tertentu seringkali tidak mengikuti pola periodic, tetapi tetap konsisten setiap tahunnya, sehingga penting untuk memasukkannya ke dalam peramalan. Memasukkan daftar hari libur ini ke dalam model dilakukan dengan mengasumsikan bahwa efek hari libur bersifat independent. Ini dilakukan dengan cara yang sama seperti seasonality yang menghasilkan matriks regressor sebagai berikut:

$$Z(t) = [1(t \in D_1), \dots, 1(t \in D_L)] \quad (15)$$

Dengan persamaan:

$$h(t) = Z(t)k \quad (16)$$

sama seperti dalam seasonality, yakni menggunakan *prior* $k \sim \text{Normal}(0, v^2)$. Sering kali penting untuk memasukkan efek *window of days* di sekitar hari libur. Prophet menyertakan parameter tambahan untuk hari-hari di sekitar hari libur, yang pada dasarnya memperlakukan hari-hari tersebut sebagai hari libur itu sendiri.

5. Optimasi Hyperparameter

Nilai prediksi yang dihasilkan dari model hybrid secara utama tergantung pada pemilihan parameter model dan parameter utama dari masing-masing model ditunjukkan pada Tabel 1. Untuk memperoleh nilai prediksi yang optimal, ketiga parameter model tersebut dioptimalkan menggunakan metode grid search dengan cross validasi. Metode ini menguji berbagai kombinasi nilai parameter model secara sistematis untuk menemukan konfigurasi yang optimal [18]. Penggunaan cross-validation penting karena membantu memastikan bahwa model tidak hanya bekerja baik pada satu set data tertentu, tetapi juga pada data yang belum pernah dilihat sebelumnya. Hal ini mengurangi risiko *overfitting* dan meningkatkan kemampuan model dalam memprediksi data baru.

Tabel 1. Hyperparameter Model DWT, SVR dan Prophet

Model	Parameter	Deskripsi Parameter	Range
DWT	Nama Wavelet	Fungsi basis wavelet	['Daubechies (db)', 'Haar', 'Coiflets', 'Symlets']
	Level Dekomposisi	Jumlah level dekomposisi	-
SVR	Kernel	Fungsi kernel	['linear', 'polinomial', 'rbf (radial basis function)', 'sigmoid', 'precomputed']

	C (Regularization Parameter)	Koefisien pinalti	['0.1', '1', '10', '100', '1000']
	γ (Gamma)	Koefisien fungsi kernel	['0.001', '0.01', '0.1', '1', '10', '100', '100']
	Epsilon	Koefisien error	['0.1', '0.2', '0.5', '1.0']
	changepoint_prior_scale	Fleksibilitas tren	[0.1,0.5]
	yearly_seasonality	Fourier order tahunan	[5,50]
Prophet	seasonality_prior_scale	Fleksibilitas musiman	[0.1,10]
	holidays_prior_scale	Fleksibilitas komponen hari libur	[0.1,10]
	seasonality_mode	Mode musiman	['additive','multiplicative']

6. Evaluasi Model

Model dievaluasi menggunakan SMAPE (Symmetric Mean Absolute Percentage Error) karena SMAPE memberikan ukuran kesalahan yang simetris dan tidak sensitif terhadap nilai-nilai ekstrem [19]. Hal ini memungkinkan SMAPE untuk memberikan gambaran yang lebih baik tentang seberapa dekat nilai prediksi dengan nilai aktual, sehingga lebih akurat dalam menilai kinerja model. SMAPE dihitung dengan rumus sebagai berikut:

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|A_t - F_t|}{(|A_t| + |F_t|)/2} \times 100\% \quad (17)$$

dengan,

A_t adalah nilai aktual pada waktu ke- t

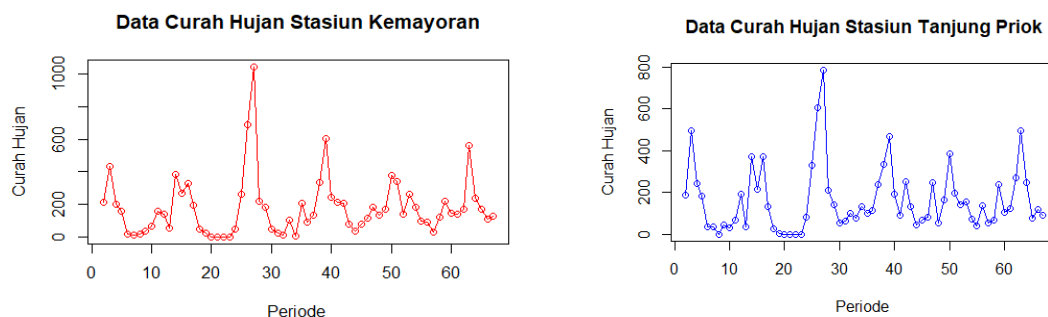
F_t adalah nilai prediksi pada waktu ke- t

n adalah jumlah periode waktu yang dievaluasi

C. Hasil dan Pembahasan

1. Data

Data yang akan digunakan dalam penelitian ini, yaitu data curah hujan bulanan dari Stasiun Meteorologi Kemayoran dan Stasiun Meteorologi Maritim Tanjung Priok seperti terlihat pada Gambar 1.



Gambar 1. Plot Data Curah Hujan Bulanan Stasiun Kemayoran dan Stasiun Tanjung Priok

Untuk mengetahui seberapa efektif model prediksi curah hujan di Jakarta, maka dilakukan evaluasi dengan metode cross validasi untuk mengukur efektivitas model prediksi curah hujan. Model yang diuji meliputi model Hybrid, SVR dan Prophet. Berikut adalah hasil rata-rata SMAPE cross validasi:

Tabel 2. Nilai SMAPE Cross Validasi

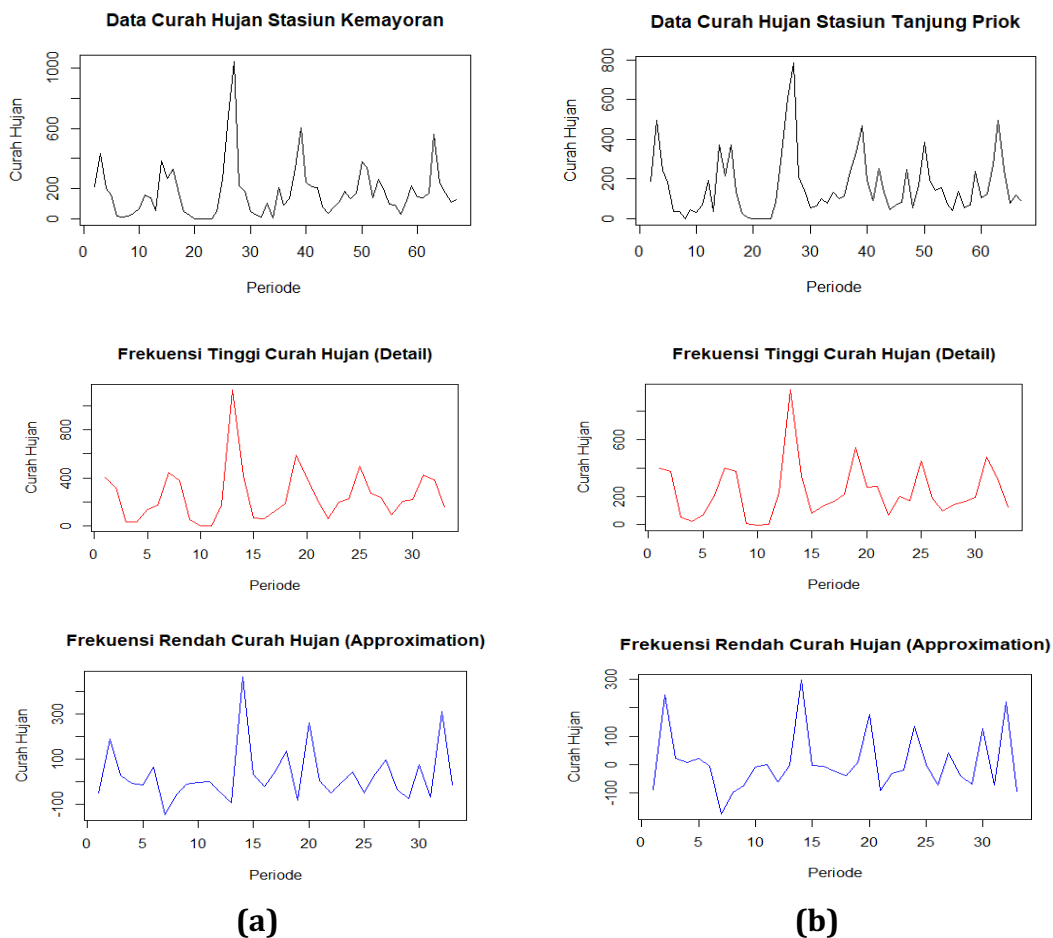
No.	Model	Rata-rata SMAPE Cross Validasi	
		Stasiun Kemayoran	Stasiun Tanjung Priok
1.	Hybrid (SVR Frekuensi Tinggi & Prophet Frekuensi Rendah)	51.11%	65.7%
2.	Hybrid (SVR Frekuensi Rendah & Prophet Frekuensi Tinggi)	96.43%	98.58%
3.	SVR	79.25%	79.41%
4.	Prophet	98.38%	100.73%

Hasil cross validasi pada Tabel 1, menunjukkan bahwa model Hybrid yang menggabungkan SVR Frekuensi Tinggi dan Prophet Frekuensi Rendah memberikan akurasi yang baik untuk Stasiun Kemayoran dan Tanjung Priok, dengan rata-rata SMAPE yaitu, 51.11% dan 65.7%. Secara keseluruhan, hasil cross validasi menunjukkan bahwa dibandingkan model lain, model dengan menggabungkan SVR frekuensi tinggi dan Prophet frekuensi rendah menghasilkan prediksi yang lebih akurat. Sangat penting untuk dicatat bahwa Stasiun Kemayoran cenderung memiliki nilai SMAPE yang lebih rendah dibandingkan dengan Stasiun Tanjung Priok. Ini menunjukkan bahwa prediksi curah hujan di Kemayoran mungkin lebih akurat selama proses cross validasi ini.

2. Model Discrete Wavelet Transform

Dengan menerapkan *Discrete Wavelet Transform*, data deret waktu berhasil dipecah menjadi beberapa sub-band frekuensi yang berbeda. Komponen frekuensi rendah (*approximation coefficients*) menggambarkan tren utama dalam data, sementara komponen frekuensi tinggi (*detail coefficients*) menangkap variasi cepat dan fluktuasi.

Penelitian ini menggunakan *Discrete Wavelet Transform* (DWT) dengan fungsi wavelet *Daubechies* untuk mendekomposisi data deret waktu. Proses dekomposisi dilakukan hingga level 1, menghasilkan dua komponen utama yaitu, komponen frekuensi tinggi (*detail*) dan komponen frekuensi rendah (*aproksimasi*). Selanjutnya komponen frekuensi tinggi (*detail*) dan komponen frekuensi rendah (*aproksimasi*) digunakan untuk memprediksi menggunakan model SVR dan model Prophet.



Gambar 2. (a) Plot Hasil Dekomposisi DWT Stasiun Kemayoran **(b)** Plot Hasil Dekomposisi DWT Stasiun Tanjung Priok

Hasil DWT yang memisahkan data curah hujan komponen frekuensi tinggi dan komponen frekuensi rendah terlihat pada Gambar 2. Komponen frekuensi tinggi memperlihatkan pola musiman yang ditandai dengan puncak curah hujan yang mengindikasikan siklus musim hujan tahunan. Meski demikian, frekuensi tinggi tetap menunjukkan kompleksitas dan variabilitas yang menggambarkan sifat cuaca suatu daerah berubah-ubah. Hal ini memberikan wawasan tentang fluktuasi curah hujan jangka pendek, yang krusial untuk pengelolaan sumber daya air, pertanian, dan strategi penanggulangan banjir. Komponen frekuensi rendah mencerminkan fluktuasi curah hujan yang lebih halus dan stabil dengan pola musiman yang cukup jelas. Frekuensi rendah memberikan gambaran tentang tren jangka panjang dalam curah hujan, yang berguna untuk memahami kondisi iklim yang lebih stabil dan perencanaan pengelolaan sumber daya air. Fluktuasi ini, meskipun lebih lembut, tetap mencerminkan variabilitas cuaca yang ada, memberikan informasi penting bagi sektor pertanian dan mitigasi bencana.

3. Model Support Vector Regression

Komponen frekuensi tinggi (detail) dan komponen frekuensi rendah (aproksimasi) yang dihasilkan dari mendekomposisi DWT selanjutnya akan diprediksi menggunakan *Support Vector Regression*. Parameter model SVR untuk

komponen frekuensi tinggi dan rendah diestimasi menggunakan metode grid search dengan cross validasi [20].

Berikut adalah estimasi parameter yang digunakan untuk memprediksi curah hujan di Jakarta:

Tabel 3. Hasil Estimasi Parameter Model SVR

Parameter	Nilai Parameter					
	Stasiun Kemayoran			Stasiun Tanjung Priok		
	Frekuensi Tinggi	Frekuensi Rendah	Tanpa DWT	Frekuensi Tinggi	Frekuensi Rendah	Tanpa DWT
Kernel	rbf	rbf	rbf	rbf	rbf	rbf
C (Regularization Parameter)	1	1	10	1	1	10
γ (Gamma)	10	1	10	10	1	10
Epsilon	0.5	0.1	0.2	0.1	0.1	0.5

Pada Tabel 3, model SVR menggunakan kernel RBF (*Radial Basis Function*) karena dapat menangani hubungan non-linear dalam data curah hujan. Nilai parameter $C = 1$ menunjukkan keseimbangan antara kompleksitas model dan training error. Nilai C yang lebih tinggi ($C = 10$) dipilih untuk data tanpa DWT, menunjukkan model lebih fokus pada mengurangi kesalahan pelatihan (training error). Nilai gamma = 10 untuk data frekuensi tinggi dan data tanpa DWT, nilai gamma yang tinggi membuat model menjadi lebih sensitif terhadap data training, sehingga model akan fokus pada pola yang spesifik, sedangkan nilai gamma = 1 untuk data frekuensi rendah membuat model lebih memperhatikan pola-pola yang lebih luas dan umum. Nilai Epsilon bervariasi antara 0,1 dan 0,5, dengan nilai yang lebih rendah pada frekuensi rendah, menunjukkan toleransi kesalahan yang lebih ketat untuk komponen ini. Estimasi parameter yang digunakan model SVR pada komponen frekuensi tinggi dan rendah menunjukkan di kedua Stasiun menunjukkan kesamaan karakteristik data curah hujan di kedua Stasiun. Perbedaan parameter untuk data tanpa DWT menunjukkan bahwa model ini disesuaikan untuk menangani kompleksitas data asli yang belum didekomposisikan oleh model DWT.

4. Model Prophet

Komponen frekuensi tinggi (detail) dan komponen frekuensi rendah (aproksimasi) diprediksi menggunakan parameter dari model Prophet. Metode grid search digunakan untuk meningkatkan kinerja keseluruhan model dengan menemukan kombinasi parameter yang optimal [21]. Penggunaan grid search dengan cross validation diharapkan akan menghasilkan prediksi yang lebih akurat dengan mengurangi overfitting dan meningkatkan generalisasi model [17].

Berikut adalah estimasi parameter yang digunakan untuk memprediksi curah hujan di Jakarta menggunakan model Prophet:

Tabel 4. Hasil Estimasi Parameter Model Prophet

Parameter	Nilai Parameter
-----------	-----------------

	Stasiun Kemayoran			Stasiun Tanjung Priok		
	Frek Tinggi	Frek Rendah	Tanpa DWT	Frek Tinggi	Frek Rendah	Tanpa DWT
changeoint_prior_scale	0.5	0.5	0.1	0.01	0.001	0.5
yearly_seasonality	1	1	1	1	1	1
seasonality_prior_scale	0.1	0.1	0.1	0.1	0.1	0.1
holidays_prior_scale	10	10	10	10	10	10
seasonality_mode	additive	additive	additive	additive	additive	additive

Pada Tabel 4, estimasi parameter model Prophet yang digunakan untuk prediksi pada komponen frekuensi tinggi, frekuensi rendah dan data tanpa DWT di Jakarta. Parameter yang dianalisis meliputi `changeoint_prior_scale`, yang mengatur fleksibilitas tren; `yearly_seasonality`, yang menetapkan order Fourier untuk komponen musiman tahunan; `seasonality_prior_scale`, yang mengontrol fleksibilitas komponen musiman; `holidays_prior_scale`, yang mengatur fleksibilitas efek hari libur; dan `seasonality_mode`, yang menentukan mode musiman. Hasil grid search menunjukkan bahwa nilai optimal untuk `changeoint_prior_scale` bervariasi antara lokasi stasiun dan jenis prediksi, sementara parameter lainnya cenderung konsisten. Variasi dalam `changeoint_prior_scale` menunjukkan bahwa data dari stasiun dan komponen frekuensi yang berbeda memiliki pola perubahan tren yang unik, sehingga model harus menyesuaikan tingkat fleksibilitasnya untuk menangkap pola tersebut secara akurat.

5. Model Hybrid

Model Hybrid merupakan model gabungan yang terdiri dari model SVR dengan model Prophet. Model Hybrid ini didapatkan dengan menjumlahkan hasil prediksi dari model SVR dan model Prophet. Berikut adalah hasil prediksi yang diperoleh dari model SVR dan model Prophet:

Tabel 5. Hasil Prediksi Curah Hujan Stasiun Kemayoran

Bulan	Data Aktual	Stasiun Kemayoran		SVR	Prophet
		Hybrid (SVR Frek Tinggi & Prophet Frek Rendah)	Hybrid (SVR Frek Rendah & Prophet Frek Tinggi)		
Januari	170.9	291.9	260.27	152.18	157.91
Februari	561.3	262.35	258.03	135.95	158.42
Maret	236	232.81	255.79	117.47	158.93
April	167.5	205.56	253.25	103.4	159.44
Mei	106.2	180.61	250.39	96.97	159.95
Juni	127.3	155.65	247.54	97.88	160.47

Hasil prediksi yang ditampilkan pada Tabel 4, menunjukkan bahwa model Hybrid (SVR Frekuensi Tinggi dan Prophet Frekuensi Tinggi) mendekati nilai aktual pada bulan Maret dengan prediksi sebesar 232.81. Selain itu, prediksi dari model tanpa DWT, yaitu SVR mendekati nilai aktual pada bulan Mei dengan prediksi

sebesar 96.97 dan hasil prediksi Prophet mendekati nilai aktual pada bulan April dengan prediksi sebesar 159.44.

Tabel 6. Hasil Prediksi Curah Hujan Stasiun Tanjung Priok

Stasiun Tanjung Priok					
Bulan	Data Aktual	Hybrid (SVR Frek Tinggi & Prophet Frek Rendah)	Hybrid (SVR Frek Rendah & Prophet Frek Tinggi)	SVR	Prophet
Januari	268.2	396.69	251.78	151.63	302.94
Februari	496.1	354.7	238.88	147.32	199.88
Maret	246.7	312.71	226	140.9	250.05
April	76.9	260.84	213.6	134.81	164.4
Mei	117.7	199.08	201.73	130.34	114.63
Juni	90.6	137.33	189.84	127.81	86.05

Hasil prediksi yang ditampilkan pada Tabel 5, menunjukkan bahwa model Prophet mendekati nilai aktual pada bulan Mei dan Juni dengan prediksi sebesar 114.63 dan 86.05. Selain itu, hasil prediksi model Hybrid (SVR Frekuensi Tinggi dan Prophet Frekuensi Rendah) mendekati nilai aktual pada bulan Juni dengan prediksi sebesar 137.33. Selanjutnya, model akan dievaluasi menggunakan metrik SMAPE untuk Stasiun Kemayoran dan Stasiun Tanjung Priok.

Tabel 7. Nilai SMAPE Model SVR, Prophet dan Hybrid

No.	Model	SMAPE	
		Stasiun Kemayoran	Stasiun Tanjung Priok
1.	Hybrid (SVR Frekuensi Tinggi & Prophet Frekuensi Rendah)	36.44%	49.46%
2.	Hybrid (SVR Frekuensi Rendah & Prophet Frekuensi Tinggi)	51.56%	50.42%
3.	SVR	47.18%	52.92%
4.	Prophet	37.88%	29.82%

Berdasarkan nilai SMAPE pada Tabel 6, model Hybrid yang menggabungkan SVR Frekuensi Tinggi dan Prophet Frekuensi Rendah memiliki kinerja terbaik di Stasiun Kemayoran dengan SMAPE sebesar 36.44%. Nilai SMAPE untuk model tanpa DWT yaitu, SVR dan Prophet masing-masing memiliki SMAPE sebesar 47.18% dan 37.88%. Hasil ini menunjukkan bahwa model Hybrid dengan DWT dan model non-Hybrid tanpa DWT memiliki akurasi yang lebih baik dibandingkan dengan model lainnya. Model Hybrid untuk Stasiun Tanjung Priok yang menggabungkan SVR Frekuensi Tinggi dan Prophet Frekuensi Rendah memiliki SMAPE sebesar 49.46%. Model tanpa DWT, yaitu SVR dan Prophet memiliki SMAPE masing-masing sebesar 52.92% dan 29.81%. Hasil ini menunjukkan bahwa model Prophet tanpa DWT memiliki kinerja yang lebih baik dibanding model Hybrid dengan DWT.

D. Simpulan

Berdasarkan hasil dari analisis yang telah dilakukan, dapat disimpulkan bahwa model Hybrid lebih unggul dibandingkan model lainnya di Stasiun Kemayoran. Hal ini terlihat dari model Hybrid yang menggabungkan SVR Frekuensi Tinggi dan Prophet Frekuensi Rendah memiliki kinerja cukup baik dengan SMAPE sebesar 36.44%. Sementara itu, di Stasiun Tanjung Priok, model non-Hybrid lebih

unggul dengan Prophet tanpa DWT yang memiliki kinerja terbaik dengan SMAPE sebesar 29.81%. Penelitian ini mengindikasikan bahwa dekomposisi wavelet tidak selalu meningkatkan akurasi prediksi. Namun, dalam data curah hujan di Jakarta model Hybrid yang menggabungkan SVR Frekuensi Rendah dan Prophet Frekuensi Tinggi memiliki SMAPE yang lebih tinggi, ini menunjukkan bahwa penggabungan model SVR Frekuensi Rendah dan Prophet Frekuensi Tinggi kurang efektif untuk karakteristik data curah hujan di wilayah Jakarta.

E. Referensi

- [1] S. Poornima, M. Pushpalatha, R. B. Jana, and L. A. Patti, "Rainfall Forecast and Drought Analysis for Recent and Forthcoming Years in India," *Water (Switzerland)*, vol. 15, no. 3, Feb. 2023, doi: 10.3390/w15030592.
- [2] K. E. Trenberth, L. Cheng, P. Jacobs, Y. Zhang, and J. Fasullo, "Hurricane Harvey Links to Ocean Heat Content and Climate Change Adaptation," *Earths Future*, vol. 6, no. 5, pp. 730–744, May 2018, doi: 10.1029/2018EF000825.
- [3] Siswanto, G. Jan van Oldenborgh, G. van der Schrier, G. Lenderink, and B. van den Hurk, "Trends in High-Daily Precipitation Events in Jakarta and the Flooding of January 2014," *Bull Am Meteorol Soc*, vol. 96, no. 12, pp. S131–S135, Dec. 2015, doi: 10.1175/bams-eee_2014_ch26.1.
- [4] A. R. As-syakur, T. Tanaka, T. Osawa, and M. S. Mahendra, "Indonesian rainfall variability observation using TRMM multi-satellite data," *Int J Remote Sens*, vol. 34, no. 21, pp. 7723–7738, 2013, doi: 10.1080/01431161.2013.826837.
- [5] M. U. M. Rao, K. C. Patra, S. K. Sasmal, A. Sharma, and G. Oliveto, "Forecasting of Rainfall across River Basins Using Soft Computing Techniques: The Case Study of the Upper Brahmani Basin (India)," *Water (Switzerland)*, vol. 15, no. 3, Feb. 2023, doi: 10.3390/w15030499.
- [6] M. Mahmudi, A. R. Hidayat, and M. Y. Wijaya, "The Effect of Ensemble Averaging Method on Rainfall Forecasting in Jakarta Using ARIMA and ARIMAX," *Mathline : Jurnal Matematika dan Pendidikan Matematika*, vol. 9, no. 2, pp. 501–512, Apr. 2024, doi: 10.31943/mathline.v9i2.608.
- [7] I. Bella Safitri, D. Urfina Zulkifli, and J. Penelitian Didaktik Matematika, "Kontinu: Jurnal Penelitian Didaktik Matematika Modeling Rainfall in Jakarta with Hybrid ARIMAX-ANN Model," vol. 8, no. 1, 2024.
- [8] H. Husain and A. Amran, "Forecasting Time Series Data Using Haar Discrete Wavelet Transformation," *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 3, pp. 484–497, May 2023, doi: 10.20956/j.v19i3.24807.
- [9] N. Asyraf, "Peramalan Curah Hujan Di Kota Medan menggunakan Metode Support Vector Regression," *Journal of Informatics and Data Science (J-IDS)*, vol. 1, no. 1, pp. 1–3, 2022.
- [10] A. Sulasikin, Y. Nugraha, J. I. Kanggrawan, and A. L. Suherman, "Monthly Rainfall Prediction Using the Facebook Prophet Model for Flood Mitigation in Central Jakarta," in *8th International Conference on ICT for Smart Society: Digital Twin for Smart Society, ICISS 2021 - Proceeding*, Institute of Electrical and Electronics Engineers Inc., Aug. 2021. doi: 10.1109/ICISS53185.2021.9532507.

- [11] D. Li *et al.*, "Prediction of Rainfall Time Series Using the Hybrid DWT-SVR-Prophet Model," *Water (Switzerland)*, vol. 15, no. 10, May 2023, doi: 10.3390/w15101935.
- [12] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2018.
- [13] M. Kaboudan, "Extended Daily Exchange Rate Forecast using Wavelet Temporal Resolutions," *New Math Nat Comput*, vol. 1, no. 1, pp. 79–107, 2005, doi: 10.1142/S1793005705000056.
- [14] Z. Liu, P. Zhou, G. Chen, and L. Guo, "Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting," *J Hydrol (Amst)*, vol. 519, pp. 2822–2831, 2014, doi: 10.1016/j.jhydrol.2014.06.050.
- [15] Y. Essam, Y. F. Huang, A. H. Birima, A. N. Ahmed, and A. El-Shafie, "Predicting suspended sediment load in Peninsular Malaysia using support vector machine and deep learning algorithms," *Sci Rep*, vol. 12, p. 302, 2022, doi: 10.1038/s41598-021-04419-w.
- [16] T. Toharudin, S. R. Pontoh, E. R. Caraka, S. Zahroh, Y. Lee, and C. R. Chen, "Employing long short-term memory and Facebook prophet model in air temperature forecasting," *Commun Stat Simul Comput*, vol. 0, no. 0, pp. 279–290, 2020, doi: 10.1080/03610918.2020.1854302.
- [17] S. J. Taylor and B. Letham, "Forecasting at Scale," *Am Stat*, vol. 72, no. 1, pp. 37–45, 2018, doi: 10.1080/00031305.2017.1380080.
- [18] D. T. Bui, P. Tsangaratos, V. T. Nguyen, N. Van Liem, and P. T. Trinh, "Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment," *Catena (Amst)*, vol. 188, May 2020, doi: 10.1016/j.catena.2019.104426.
- [19] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput Sci*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.
- [20] L. Villalobos-Arias, C. Quesada-López, J. Guevara-Coto, A. Martínez, and M. Jenkins, "Evaluating hyper-parameter tuning using random search in support vector machines for software effort estimation," in *PROMISE 2020 - Proceedings of the 16th ACM International Conference on Predictive Models and Data Analytics in Software Engineering, Co-located with ESEC/FSE 2020*, Association for Computing Machinery, Inc, Nov. 2020, pp. 31–40. doi: 10.1145/3416508.3417121.
- [21] L. Menculini *et al.*, "Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices," *Forecasting*, vol. 3, no. 3, pp. 644–662, Sep. 2021, doi: 10.3390/forecast3030040.