

Analisis Perbandingan Support Vector Regression dan K-Nearest Neighbor Regression untuk Prediksi Produksi TBS Bulanan (Studi Kasus: CV. Bayu Sukma Jaya)

Bayu Sukma Sani*, Muhammad Iqbal, Darmeli Nasution, Muhammad Syahputra Novelan, Zulham Sitorus

Magister Teknologi Informasi, Pascasarjana, Universitas Pembangunan Panca Budi, Indonesia

Email: ¹*bayuzkiki@gmail.com, ²muhammadiqbal@dosen.pancabudi.ac.id, ³darmelinasution@dosen.pancabudi.ac.id, ⁴putranovelan@dosen.pancabudi.ac.id, ⁵zulhamsitorus@dosen.pancabudi.ac.id

Email Penulis Korespondensi: bayuzkiki@gmail.com*

Submitted: 12/06/2026; Accepted: 29/06/2026; Published: 30/06/2026

Abstrak—Produksi tandan buah segar (TBS) kelapa sawit berubah setiap bulan karena kondisi tanaman, lingkungan, iklim, dan kegiatan perkebunan lainnya. Perencanaan tenaga kerja, transportasi, pengolahan, dan pencapaian target produksi perusahaan dapat dipengaruhi oleh fluktuasi tersebut. Tujuan dari penelitian ini adalah untuk membandingkan kinerja Support Vector Regression (SVR) dan K-Nearest Neighbor Regression (KNN) algoritma dalam memprediksi produksi TBS bulanan pada CV. Bayu Sukma Jaya. Data penelitian digunakan dari periode tahun awal hingga tahun akhir, dengan variabel prediktor seperti curah hujan, hari hujan, luas panen, umur tanaman, pemupukan, atau variabel lain yang digunakan. Pengumpulan, pembersihan, analisis awal, normalisasi, pembagian data latih dan uji, pembentukan model, pengujian parameter, dan evaluasi kinerja model adalah bagian dari proses penelitian. Untuk mendapatkan konfigurasi terbaik untuk setiap algoritma, optimalisasi parameter dilakukan. Hasil pengujian digunakan untuk mengevaluasi kinerja model SVR/KNN Regression. Dengan nilai MAE sebesar nilai, RMSE sebesar nilai, MAPE sebesar nilai, dan R2 sebesar nilai, nilai kesalahan absolut (MAE) sebesar nilai, dan koefisien determinasi (R2) sebesar nilai. Nilai kesalahan absolut yang lebih rendah dan koefisien determinasi yang lebih tinggi menunjukkan bahwa model tersebut memiliki tingkat kesalahan absolut yang lebih tinggi. Hasil penelitian menunjukkan bahwa algoritma SVR/KNN Regression dapat digunakan sebagai metode untuk memprediksi produksi TBS pada CV. Bayu Sukma Jaya. Hasil prediksi ini dapat membantu perusahaan membuat keputusan tentang perencanaan produksi, kebutuhan tenaga kerja, transportasi hasil panen, dan pengambilan keputusan operasional dengan lebih akurat.

Kata Kunci: *K-Nearest Neighbor Regression; Kelapa Sawit; Prediksi Produksi; Support Vector Regression; TBS.*

Abstract— The production of fresh fruit bunches (FFB) of oil palm changes every month due to the condition of the plants, environment, climate, and other plantation activities. Labour, transportation, processing, and achievement of the company's production targets can be affected by these fluctuations. The objective of this research is to compare the performance of the Support Vector Regression (SVR) and K-Nearest Neighbour Regression (KNN) algorithms in predicting monthly CPO production at CV. Bayu Sukma Jaya. The research data used spans the period from start year to end year, with predictor variables such as rainfall, rainy days, harvest area, plant age, fertilisation, or other variables used. Collection, cleaning, initial analysis, normalisation, splitting of training and test data, model formation, parameter testing, and model performance evaluation are part of the research process. To obtain the best configuration for each algorithm, parameter optimisation is performed. The test results are used to evaluate the performance of the [SVR/KNN Regression] model. With an MAE value of value, RMSE of value, MAPE of value, and R2 of value, the absolute error (MAE) value of value, and the coefficient of determination (R2) of value. A lower absolute error value and a higher coefficient of determination indicate that the model has a higher absolute error rate. The research results show that the SVR/KNN Regression algorithm can be used as a method to predict CPO production at CV. Bayu Sukma Jaya. These prediction results can help the company make decisions about production planning, labour needs, transportation of harvests, and operational decision-making more accurately.

Keywords: *K-Nearest Neighbour Regression; Palm Oil; Production Prediction; Support Vector Regression; TBS*

1. PENDAHULUAN

Salah satu komoditas perkebunan yang sangat penting untuk perekonomian Indonesia adalah kelapa sawit. Komoditas ini tidak hanya meningkatkan sektor industri pengolahan minyak nabati, tetapi juga meningkatkan pendapatan lokal dan aktivitas ekonomi perusahaan perkebunan. Menurut Pusat Statistik, data tentang kelapa sawit nasional disusun berdasarkan luas areal, produksi, status pengusahaan, ekspor, dan impor. Ini menunjukkan bahwa kelapa sawit adalah salah satu komoditas perkebunan yang paling penting di Indonesia [1].

Produksi Tandan Buah Segar (TBS) adalah faktor utama yang menentukan keberhasilan produksi dalam kegiatan operasional perkebunan sawit. Perencanaan panen, kebutuhan tenaga kerja, kapasitas angkut, biaya operasional, dan estimasi pendapatan perusahaan semua dipengaruhi oleh jumlah TBS yang dihasilkan setiap bulan. Namun, berbagai faktor, seperti umur tanaman, kondisi cuaca, curah hujan, pemupukan, tingkat kesuburan tanah, serangan hama, dan kesiapan buah untuk dipanen, cenderung memengaruhi produksi TBS bulanan.

Produksi kelapa sawit tidak selalu stabil setiap bulan, [2] karena berbagai faktor, termasuk iklim, kesuburan tanah, curah hujan yang tidak menentu, dan kondisi tandan buah yang belum siap panen.

Perusahaan perkebunan menghadapi tantangan dengan fluktuasi produksi TBS tersebut, terutama bagi masyarakat yang memiliki lahan perkebunan sawit. Dalam membuat perencanaan produksi bulanan untuk TBS hasil panen dari Masyarakat, CV. Bayu Sukma Jaya menemukan bahwa ada perubahan dalam proses hasil panen sawit dari Masyarakat. Hasil penjualan TBS dari Masyarakat ke CV. Bayu Sukma Jaya menunjukkan proses ini. Jika masyarakat hanya bergantung pada perkiraan manual yang dibuat berdasarkan pengalaman, hasil prediksi mungkin kurang akurat, terutama jika pola produksi berubah setiap bulan. Jumlah tenaga kerja, jadwal panen, kebutuhan transportasi, dan target produksi yang telah direncanakan dapat berbeda karena prediksi yang tidak akurat. Akibatnya, masyarakat membutuhkan metode analitis berbasis data historis agar prediksi produksi TBS bulanan dapat dilakukan secara lebih terukur dan tidak bias.

Untuk memprediksi data numerik, metode regresi K-Nearest Neighbor, juga dikenal sebagai KNN Regression, dapat digunakan selain SVR. Metode ini bekerja dengan mengumpulkan sejumlah data terdekat berdasarkan kemiripan pola, kemudian menggunakan nilai tetangga terdekat sebagai dasar untuk memprediksi. Metode ini cukup sederhana dan mudah dipahami, dan dapat digunakan pada data dengan pola kedekatan antarobservasi. [3] memprediksi hasil kelapa sawit berdasarkan faktor lingkungan dengan menggunakan regresi K-Nearest Neighbor. Ini menunjukkan bahwa pendekatan berbasis kedekatan data dapat digunakan untuk memprediksi hasil tanaman kelapa sawit. Hasil penelitian menunjukkan bahwa algoritma Random Forest sedikit lebih baik daripada KNN. Dengan nilai akurasi 0,6023, presisi 0,4827, recall 0,4177, skor F1 0,4479, dan AUC 0,5681, KNN memperoleh nilai akurasi 0,5990, presisi 0,4771, recall 0,4006, dan skor F1 0,4355 [4].

Metode SVR telah diterapkan untuk memprediksi produksi kelapa sawit dalam beberapa penelitian sebelumnya. [2] menggunakan SVR untuk memprediksi produktivitas dan produksi kelapa sawit, dan menemukan bahwa kernel RBF menghasilkan prediksi yang mendekati nilai aktual. [5] membandingkan kernel SVR untuk memprediksi produksi kelapa sawit, dan menunjukkan bahwa pemilihan kernel mempengaruhi performa prediksi. Namun, penelitian yang secara khusus membandingkan SVR dan KNN Regression pada produksi TBS bulanan di CV. Bayu Sukma Jaya masih perlu dilakukan. Ini terutama untuk menentukan metode mana yang paling sesuai dengan pola data perusahaan. [6] Hasil dari metode K-Nearest Neighbor Test Hasil analisis metode K-NN menunjukkan persentase yang lebih besar dari nilai K 1 hingga 10. [7], Data yang digunakan berasal dari siswa SMA Negeri 1 Tanjung Pura. Algoritma K-Nearest Neighbor akan digunakan untuk mengklasifikasikan data ini untuk menentukan apakah seseorang kecanduan belanja online berdasarkan tingkat kecanduannya. Hasilnya akan menunjukkan apakah tingkat kecanduan belanja online siswa tersebut rendah atau tinggi.

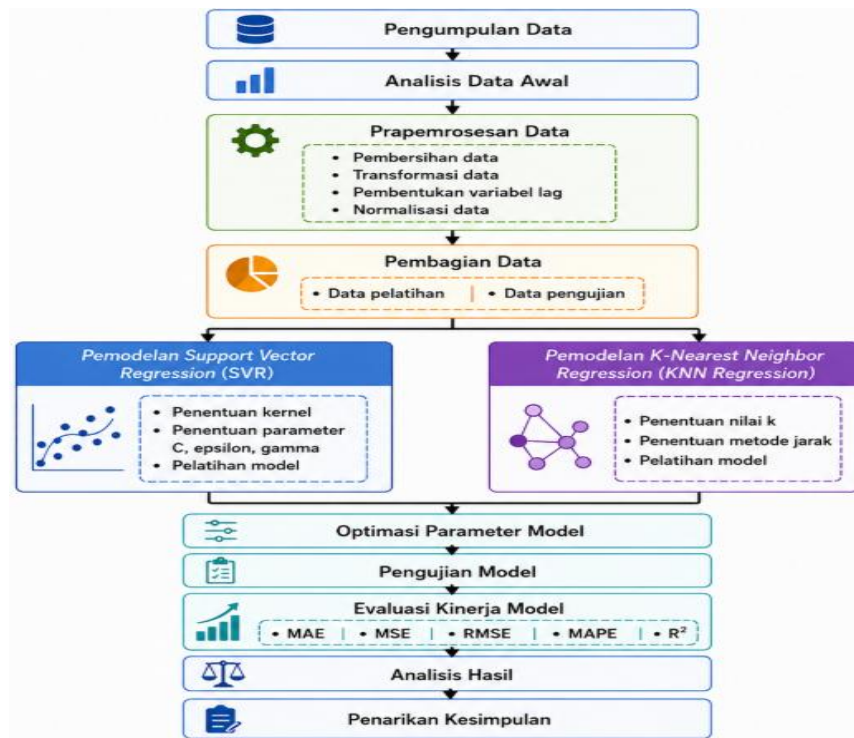
Berdasarkan ukuran evaluasi seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Mean Absolute Percentage Error (MAPE), penelitian ini diharapkan dapat menghasilkan model prediksi produksi TBS bulanan yang lebih akurat. Penelitian ini diharapkan dapat membantu masyarakat dalam perencanaan panen, estimasi produksi, kebutuhan operasional, pengambilan keputusan berbasis data, dan CV. Bayu Sukma Jaya sebagai rekanan masyarakat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode komparatif dan pendekatan kuantitatif digunakan dalam penelitian ini untuk mengevaluasi kinerja algoritma Support Vector Regression (SVR) dan K-Nearest Neighbor Regression (KNN Regression) dalam memprediksi produksi tandan buah segar (TBS) bulanan pada CV. Bayu Sukma Jaya. Data sekunder yang digunakan adalah deret waktu produksi TBS bulanan yang diperoleh dari dokumentasi perusahaan. Jumlah produksi TBS digunakan sebagai variabel target dalam penelitian. Di sisi lain, data sejarah produksi dan variabel lain yang tersedia, seperti luas panen, curah hujan, jumlah hari kerja, atau faktor operasional, digunakan sebagai variabel prediktor. Untuk mencegah kebocoran data, pengolahan data mencakup pemeriksaan data kosong, penanganan data pencilan, transformasi, normalisasi, dan pembagian data menjadi data latih dan data uji dalam urutan waktu tertentu.

Penelitian ini melakukan pemrosesan data sesuai data set yang ada dari sumber usaha dibidang produksi TBS. Setelah dibagi menjadi data pelatihan dan pengujian, data produksi TBS dimasukkan ke dalam model regresi SVR dan KNN. Hasil prediksi kedua model kemudian dibandingkan berdasarkan nilai kesalahannya; model dengan nilai kesalahan paling rendah dan nilai R2 tertinggi dipilih untuk memprediksi produksi TBS bulanan di CV. Bayu Sukma Jaya.



Gambar 1. Kerangka Penelitian

Metode penelitian dimulai dengan pengumpulan data produksi TBS bulanan pada CV. Bayu Sukma Jaya. Setelah itu, data awal dianalisis untuk mengidentifikasi ciri, pola, dan kualitas data yang tersedia. Setelah itu, proses prapemrosesan data dilakukan. Ini mencakup pembersihan, transformasi, pembentukan variabel lag, dan normalisasi data agar siap digunakan dalam proses pemodelan. Data kemudian dibagi menjadi data pengujian dan pelatihan, dan keduanya diproses menggunakan regresi vektor pendukung dan regresi K-Nearest Neighbor. Penentuan parameter, pelatihan model, dan optimasi parameter dilakukan pada masing-masing model untuk mendapatkan konfigurasi optimal. Selanjutnya, model yang telah dibuat diuji dan dievaluasi menggunakan metrik MAE, MSE, RMSE, MAPE, dan R^2 . Hasilnya dianalisis untuk mengetahui perbedaan kinerja kedua algoritma. Dengan demikian, dapat diputuskan bahwa model yang paling akurat untuk memprediksi produksi TBS bulanan adalah yang terbaik.

2.2 Pengertian Kelapa Sawit

Kelapa sawit (*Elaeis guineensis*) adalah salah satu tanaman perkebunan yang paling penting di dunia, terutama karena kemampuannya menghasilkan minyak sawit, yang merupakan bahan baku utama dalam industri makanan, kosmetik, dan biodiesel. Minyak sawit dikenal karena produktivitasnya yang tinggi dan biaya produksi yang relatif rendah dibandingkan dengan minyak nabati lainnya, menjadikannya pilihan utama dalam berbagai aplikasi industri [8]. Selain itu, isu sosial seperti konflik lahan dan kondisi kerja juga menjadi perhatian. Tanaman ini memiliki siklus hidup yang panjang dan dapat berproduksi hingga puluhan tahun, menjadikannya salah satu tanaman perkebunan yang paling bernilai ekonomis di dunia [9]. Oleh karena itu, praktik budidaya yang berkelanjutan dan sertifikasi seperti RSPO (Roundtable on Sustainable Palm Oil) semakin diutamakan untuk memastikan bahwa produksi minyak sawit dilakukan dengan cara yang tidak merusak lingkungan dan menghormati hak-hak sosial masyarakat lokal.

2.3 Tandan Buah Segar (TBS)

Dalam industri pengolahan kelapa sawit, Tandan Buah Segar (TBS) merupakan bahan baku utama yang sangat menentukan kualitas dan kuantitas produk akhir. Kualitas TBS secara langsung memengaruhi rendemen minyak sawit mentah (CPO) yang dihasilkan. Oleh karena itu, pengelolaan TBS secara efektif dan efisien menjadi faktor kunci dalam menjaga kontinuitas proses produksi dan mengoptimalkan biaya operasional di pabrik kelapa sawit [10]. TBS yang belum matang (underripe) cenderung menghasilkan rendemen rendah, sementara buah yang terlalu matang (overripe) dapat menyebabkan peningkatan kadar asam lemak bebas (FFA) yang berdampak negatif terhadap mutu minyak dan meningkatkan biaya pemurnian.

Setelah proses panen, penanganan pascapanen yang tepat menjadi hal yang tidak kalah penting. TBS harus diproses atau dikirim ke pabrik dalam waktu tidak lebih dari 24 jam untuk menghindari penurunan mutu akibat proses fermentasi. Buah yang terlalu lama disimpan berisiko mengalami peningkatan FFA karena adanya

aktivitas enzim dan mikroorganisme. Selain itu, kerusakan mekanis selama pengangkutan seperti memar atau pecahnya buah juga dapat mempercepat kerusakan mutu. Penanganan pascapanen yang baik dapat mencegah penurunan kualitas akibat fermentasi atau kerusakan mekanis [11].

2.4 Data Mining

Data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. [12] data mining juga didefinisikan sebagai bagian dari proses penggalian pengetahuan dalam database yang sering disebut dengan istilah Knowledge Discovery in Database (KDD). Teknik data mining didukung oleh tiga teknologi yaitu pengumpulan data secara besar, multiprocessor pada komputer dan algoritma data mining. Tugas dari data mining yaitu Deskriptif – menemukan gambaran pola yang menarik dari data dan Prediktif - memprediksi perilaku dari model berdasarkan data yang ada.

2.5 Machine Learning

Machine Learning merupakan studi untuk membuat komputer dapat membuat keputusan secara otomatis menggunakan berbagai algoritma yang dapat ditingkatkan akurasi melalui proses pembelajaran dengan data history. Algoritma machine learning merupakan metode yang digunakan untuk menemukan pola dan untuk membuat model. Machine learning adalah metode untuk membuat program yang dapat belajar dari data yang tersedia dari berbagai sumber yang dapat dipercaya untuk digunakan [13]. Algoritma machine learning dalam klasifikasi menggunakan dataset yang tersedia untuk umum. Tujuan utama adalah untuk mengidentifikasi algoritma yang paling akurat dan efisien secara komputasi untuk penerapan praktis di lingkungan dunia nyata [14].

2.6 Algoritma Support Vector Regression

Data mining adalah proses yang digunakan untuk menemukan pola dan informasi yang berguna dari kumpulan data besar. Dalam konteks penelitian ini, data mining berperan penting dalam menganalisis data hasil panen ayam broiler untuk memprediksi kebutuhan produksi. Metode yang umum digunakan dalam data mining meliputi asosiasi, klasifikasi, regresi, dan clustering. Di antara metode tersebut, regresi, khususnya Support Vector Regression (SVR), telah terbukti efektif dalam menangani data non-linear dan kompleks [15].

SVR adalah teknik yang dikembangkan dari Support Vector Machine (SVM), yang awalnya dirancang untuk masalah klasifikasi. SVR berfungsi untuk memodelkan hubungan antara variabel independen dan dependen, memungkinkan prediksi nilai numerik berdasarkan data historis (Putri et al., n.d.) SVR memiliki keunggulan dalam menangani data multivariat dan dapat memberikan hasil yang akurat meskipun terdapat outlier dalam dataset [16]. Penelitian sebelumnya menunjukkan bahwa SVR dapat digunakan untuk berbagai aplikasi, termasuk prediksi hasil panen dan analisis harga komoditas.

$$f(x) = \sum_{i=1}^N \alpha_i K(x_i, x) + b \quad (1)$$

$$\min_{w, b, \epsilon} \left(\frac{1}{2} \|w\|^2 + c \sum_{i=1}^N \epsilon_i \right) \quad (2)$$

2.7 Algoritma K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan metode yang biasa digunakan pada klasifikasi data. Algoritma ini digunakan untuk mengklasifikasikan terhadap objek berdasarkan data pembelajaran yang jarak tetangganya paling dekat atau memiliki nilai selisih yang kecil dengan objek tersebut. KNN merupakan suatu metode yang menggunakan algoritma supervised dengan hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Penentuan banyaknya data training dan data testing dapat menggunakan dua skema meliputi: (1) pengambilan sampel acak tanpa pengembalian dengan mengabaikan proporsi kelas, atau (2) pengambilan sampel secara acak tanpa pengembalian dengan mempertimbangkan proporsi kelas. Pada algoritma KNN, kedekatan antara data training dan data testing menjadi faktor utama dalam proses klasifikasi maupun prediksi. Ukuran kedekatan ini umumnya dihitung menggunakan matrik jarak, di mana jarak Euclidean adalah yang paling umum digunakan. Rumus adalah sebagai berikut:

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

2.8 Machine Learning

Regresi K-Nearest Neighbor (KNN Regression) merupakan metode nonparametrik yang digunakan untuk memprediksi nilai kontinu dari suatu data baru. Prediksi didasarkan pada rata-rata nilai target dari sejumlah k tetangga terdekat dalam data pelatihan. Metode ini tidak membentuk model eksplisit pada saat pelatihan,

melainkan menyimpan seluruh data pelatihan dan melakukan prediksi langsung saat data baru diberikan (instance-based learning atau lazy learning).

$$y = \frac{1}{k} \sum_{i=1}^k y_i \quad (4)$$

2.9 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) adalah metrik evaluasi untuk mengukur perbedaan antara nilai prediksi yang dihasilkan oleh suatu model dengan nilai aktual. RMSE berfungsi untuk mengukur secara kuantitatif besarnya kesalahan prediksi dalam satuan yang sama dengan data aktual, sehingga memudahkan interpretasi tingkat akurasi model. RMSE memiliki akurasi yang tinggi dalam melakukan pengukuran dengan probabilitas hingga 50% pada algoritma yang digunakan untuk memprediksi data, berikut persamaan rumus:

$$RMSE = \sqrt{\sum \frac{(y' - y)^2}{n}} \quad (5)$$

2.10) Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah metrik evaluasi yang digunakan untuk mengukur rata-rata kesalahan absolut antara nilai aktual dan nilai prediksi oleh model. Semakin kecil nilai MAE, semakin akurat model tersebut dalam melakukan prediksi. MAE memberikan gambaran yang sederhana namun efektif tentang seberapa besar kesalahan prediksi dalam satuan yang sama dengan data aslinya. Berikut persamaan untuk menghitung nilai MAE:

$$MEA = \frac{1}{n} \sum |y_i - y'_i| \quad (6)$$

2.9 Variabel Bebas dan Fitur Temporal

Variabel bebas merupakan seluruh atribut yang digunakan sebagai masukan model untuk memprediksi produksi TBS bulanan. Variabel tersebut terdiri atas variabel dasar dan fitur temporal hasil rekayasa data. Variabel dasar meliputi luas tanaman menghasilkan, umur tanaman, dan periode bulan. Luas tanaman menghasilkan dinyatakan dalam hektare dan menunjukkan kapasitas lahan produktif pada setiap desa. Umur tanaman dinyatakan dalam tahun dan diperlakukan sebagai karakteristik periode serta tidak digunakan untuk menyimpulkan pengaruh kausal umur tanaman secara terpisah dari efek tahun.

$$\text{MonthSin}_t = \sin\left(\frac{2\pi m_t}{12}\right) \quad (7)$$

$$\text{MonthCos}_t = \cos\left(\frac{2\pi m_t}{12}\right) \quad (8)$$

Fitur temporal dibentuk dari riwayat produksi sebelum periode target. Fitur *lag* merepresentasikan produksi pada satu sampai tiga bulan sebelumnya dan dinotasikan.

$$\text{Lag}_k(i, t) = P_{i,t-k}, \quad k = 1, 2, 3 \quad (9)$$

$$\text{MA3}_{i,t} = \frac{P_{i,t-1} + P_{i,t-2} + P_{i,t-3}}{3} \quad (10)$$

$$\text{SD3}_{i,t} = \sqrt{\frac{\sum_{j=1}^3 (P_{i,t-j} - \text{MA3}_{i,t})^2}{3}} \quad (11)$$

3. HASIL DAN PEMBAHASAN

3.1 Skenario Percobaan

Skenario percobaan disusun untuk mengukur kontribusi setiap kelompok fitur terhadap kinerja Support Vector Regression dan K-Nearest Neighbor Regression dalam memprediksi produksi TBS bulanan. Pengujian dilakukan secara bertahap, mulai dari penggunaan variabel dasar hingga penambahan fitur temporal yang berasal dari riwayat produksi. Skenario pertama menggunakan fitur dasar berupa luas tanaman menghasilkan, umur tanaman, serta informasi bulan. Skenario berikutnya menambahkan fitur produksi bulan sebelumnya dalam bentuk (Lag1), (Lag2), dan (Lag3).

Tabel 1. Rancangan Skenario Percobaan

Skenario	Kelompok fitur	Tujuan
F0	Luas tanaman menghasilkan dan umur tanaman	Menjadi skenario dasar untuk mengukur kemampuan prediksi model menggunakan karakteristik utama perkebunan tanpa fitur temporal.
F1	F0 + MonthSin + MonthCos	Menguji kontribusi pola musiman bulanan terhadap peningkatan kinerja prediksi.
F2	F1 + Lag1	Menguji pengaruh produksi satu bulan sebelumnya terhadap produksi pada bulan target.
F3	F1 + Lag1 + Lag2 + Lag3	Menguji kemampuan model memanfaatkan riwayat produksi jangka pendek selama tiga bulan sebelumnya.
F4	F3 + MA3 + Growth + SD3	Menguji kontribusi gabungan informasi tren, perubahan, dan volatilitas produksi jangka pendek terhadap kinerja model.

Setelah itu, eksperimen dilanjutkan dengan penambahan rata-rata bergerak tiga bulan, pertumbuhan produksi, dan simpangan baku produksi sebagai representasi tren dan volatilitas jangka pendek. Rancangan skenario percobaan.

Tabel 2. Data Set

Tha	Bln	Indeks Bulan	Curah Hujan mm	Hari Hujan	Suhu Rata2 C	Kelembapan n per	Luas Tanaman ha	Umur Tanaman tahun	Pemupukan kg per ha	Tenaga Kerja Panen	Produksi TBS Lag1 ton	Produksi TBS Lag2 ton	Produksi TBS Lag3 ton	Moving Avg 3 ton	Data Split	Pro TBS
2020	1	1	255.7	21	26.8	83.6	117.9	6	49	21					Train	2
2020	2	2	243.5	21	27.3	84	118.5	6.1	63.4	19	214.5				Train	2
2020	3	3	237.4	22	27.6	81.2	118.4	6.2	57.2	20	212.2	214.5			Train	2
2020	4	4	236.2	19	27.9	81.8	118.6	6.2	50.5	22	240.1	212.2	214.5	222.3666 667	Train	2
2020	5	5	101.4	7	27.7	77.2	118.5	6.3	50.1	23	238.3	240.1	212.2	230.2	Train	2
2020	6	6	147.4	12	27.8	79.1	118.5	6.4	56.4	23	243.8	238.3	240.1	240.7333 333	Train	2
2020	7	7	111.2	8	27.4	76.9	117.8	6.5	64.3	20	280.4	243.8	238.3	254.1666 667	Train	2
2020	8	8	143.1	15	27.3	78.9	119.1	6.6	62.6	19	262.9	280.4	243.8	262.3666 667	Train	2
2020	9	9	168.6	11	26.9	78.9	117.7	6.7	65.9	22	287.1	262.9	280.4	276.8	Train	3
2020	10	10	198.6	15	26.5	78	118.8	6.8	52.7	21	309.3	287.1	262.9	286.4333 333	Train	2
2020	11	11	208.3	17	26	77.9	117.5	6.8	56.7	23	284	309.3	287.1	293.4666 667	Train	2
2020	12	12	267.8	23	26.9	84.5	118.2	6.9	51	23	260.3	284	309.3	284.5333 333	Train	2
2021	1	13	296.8	25	27.3	83.3	121.5	7	55.5	20	237.2	260.3	284	260.5	Train	2
2021	2	14	245.5	23	27.4	83.1	118.3	7.1	65.8	22	215.4	237.2	260.3	237.6333 333	Train	2
2021	3	15	211.2	18	28	78.1	119.9	7.2	68.4	20	226.8	215.4	237.2	226.4666 667	Train	2
2021	4	16	212.9	15	27.9	77.9	120.9	7.2	55	26	259.2	226.8	215.4	233.8	Train	2
2021	5	17	221	16	27.9	81.3	121.6	7.3	48.3	24	279.2	259.2	226.8	255.0666 667	Train	2
2021	6	18	196	18	27.8	78.5	119.2	7.4	55.8	21	288.2	279.2	259.2	275.5333 333	Train	3
2021	7	19	111.7	10	27.2	75.1	120.4	7.5	58.3	25	313.9	288.2	279.2	293.7666 667	Train	2
2021	8	20	86.5	8	27.3	75.8	119.6	7.6	67.5	22	289.2	313.9	288.2	297.1	Train	2
2021	9	21	178.2	16	26.9	83.1	120.4	7.7	61.8	22	292.6	289.2	313.9	298.5666 667	Train	3
2021	10	22	183.6	16	26.9	78	118.8	7.8	56.2	22	323.7	292.6	289.2	301.8333 333	Train	2
2021	11	23	256.1	23	26.3	83.2	120	7.8	49.1	22	289.7	323.7	292.6	302	Train	2
2021	12	24	237.6	21	26.7	80.2	120.5	7.9	59.1	20	258.6	289.7	323.7	290.6666 667	Train	2

3.2 Hasil

Hasil pengolahan data produksi tandan buah segar (TBS) bulanan CV. Bayu Sukma Jaya menunjukkan bahwa jumlah produksi berubah selama setiap periode. Setelah melalui tahap pembersihan, normalisasi, dan pembagian data berdasarkan urutan waktu, model Support Vector Regression dan K-Nearest Neighbor Regression dibangun dari data tersebut. Meskipun kedua model memiliki kemampuan untuk mengikuti pola perubahan produksi TBS, mereka menghasilkan tingkat kedekatan prediksi yang berbeda terhadap data aktual. Ini terlihat dari perbedaan antara nilai hasil prediksi dan nilai produksi pada data pengujian, terutama selama periode di mana produksi meningkat atau menurun secara signifikan.

Beberapa variabel historis, seperti produksi bulan sebelumnya, juga digunakan sebagai input untuk membantu model mengenali pola perubahan produksi TBS.

Tabel 3. Rekapitulasi data produksi TBS berdasarkan tahun

Tahun	Jumlah desa	Luas tanaman menghasilkan	Produksi TBS	Produktivitas tertimbang
2023	16	634 ha	14.470,55 ton	22,82 ton/ha/tahun
2024	16	634 ha	15.174,50 ton	23,93 ton/ha/tahun
2025	16	634 ha	15.060,90 ton	23,76 ton/ha/tahun
Total	-	-	44.705,95 ton	-

Produksi TBS meningkat sebesar 4,86% dari 2023 ke 2024. Pada 2025, produksi mengalami penurunan sebesar 0,75% dibandingkan 2024. Meskipun demikian, produksi 2025 masih lebih tinggi 4,08% dibandingkan produksi 2023. Nilai produksi bulanan terendah dalam dataset adalah 12, 31 ton yang tercatat di Desa Putat pada Februari 2023. Produksi bulanan tertinggi adalah 173, 23 ton yang tercatat di Desa Sekeladi pada September 2024. Perbedaan tersebut menunjukkan adanya variasi skala produksi antardesa yang terutama berkaitan dengan perbedaan luas tanaman menghasilkan.

Dataset penelitian terdiri atas 576 observasi desa-bulan yang berasal dari 16 desa selama periode Januari 2023 sampai Desember 2025. Total produksi TBS yang tercatat selama tiga tahun adalah 44.705,95 ton. Distribusi observasi pada setiap tahun bersifat seimbang, yaitu: 16 desa × 12 bulan = 192 observasi per tahun. Secara keseluruhan, rata-rata produksi setiap observasi desa-bulan adalah 77,61 ton dengan median sebesar 67,80 ton. Nilai rata-rata yang lebih tinggi daripada median menunjukkan bahwa distribusi produksi tidak seimbang. Kondisi tersebut disebabkan oleh adanya beberapa desa dengan luas tanaman menghasilkan dan volume produksi yang jauh lebih besar dibandingkan desa lainnya. Simpangan baku produksi keseluruhan sebesar 45,93 ton menunjukkan adanya variasi produksi yang cukup tinggi antardesa. Produksi bulanan terendah adalah 12,31 ton, sedangkan produksi bulanan tertinggi mencapai 173,23 ton. Statistik deskriptif produksi TBS berdasarkan tahun ditampilkan pada Tabel 4.

Tabel 4. Statistik Deskriptif Produksi TBS Bulanan

Tahun	n	Total produksi	Rata-rata	Median	Simpangan baku	Minimum	Maksimum	CV
2023	192	14.470,55 ton	75,37	65,42	44,77	12,31	167,08	59,41%
2024	192	15.174,50 ton	79,03	68,74	46,77	12,88	173,23	59,18%
2025	192	15.060,90 ton	78,44	69,56	46,36	12,85	169,90	59,10%
Keseluruhan	576	44.705,95 ton	77,61	67,80	45,93	12,31	173,23	59,17%

Dataset sebanyak 576 observasi tersebut selanjutnya digunakan sebagai dasar pembentukan fitur temporal. Pada tahap rekayasa fitur, sejumlah observasi awal pada setiap desa tidak memiliki nilai lag karena belum tersedia data produksi bulan sebelumnya.

3.3 Perbandingan Produksi TBS Antara Tahun dan Antara Desa

Total produksi TBS pada 2023 adalah 14.470,55 ton. Produksi meningkat menjadi 15.174, 50ton pada 2024 atau tumbuh sebesar 4,86%. Pada 2025, produksi menurun menjadi 15.060, 90ton atau berkurang sebesar 0,75% dibandingkan 2024. Perubahan produksi tahunan dihitung menggunakan persamaan berikut:

$$Pertumbuhan_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100\% \quad (12)$$

Tabel 5. Ringkasan Karakteristik Produksi Berdasarkan Desa

Desa	Luas TM	Rata-rata produksi tahunan	Rata-rata produktivitas	Pertumbuhan 2023–2025	CV produksi bulanan
Sekeladi	85 ha	1.763,75 ton	20,75 ton/ha	3,21%	10,64%
Sekeladi Hilir	73 ha	1.490,42 ton	20,42 ton/ha	1,50%	10,92%
Sedinginan	68 ha	1.726,07 ton	25,38 ton/ha	4,45%	10,96%
Sintong Makmur	65 ha	1.763,67 ton	27,13 ton/ha	4,73%	11,20%
Menggala Sempurna	58 ha	1.390,07 ton	23,97 ton/ha	4,09%	9,68%
Teluk Berembun	52 ha	1.176,07 ton	22,62 ton/ha	3,90%	10,52%
Teluk Mega	40 ha	918,00 ton	22,95 ton/ha	4,93%	10,93%
Ujung Tanjung	37 ha	910,82 ton	24,62 ton/ha	4,40%	10,60%
Mumugo	31 ha	672,18 ton	21,68 ton/ha	4,07%	10,48%
Sintong Bakti	30 ha	797,50 ton	26,58 ton/ha	6,43%	10,57%
Banjar XII	26 ha	675,13 ton	25,97 ton/ha	3,78%	10,66%
Sintong	18 ha	469,20 ton	26,07 ton/ha	4,94%	11,33%
Sintong Pusaka	17 ha	377,68 ton	22,22 ton/ha	3,98%	11,53%
Menggala Sakti	15 ha	309,00 ton	20,60 ton/ha	4,98%	10,19%
Rantau Bais	11 ha	276,83 ton	25,17 ton/ha	4,55%	11,86%
Putat	8 ha	185,60 ton	23,20 ton/ha	5,79%	11,10%

Sekeladi dan Sintong Makmur memiliki rata-rata produksi tahunan tertinggi, yaitu masing-masing sebesar 1.763, 75 ton dan 1.763,67 ton. Meskipun total produksinya hampir sama, kedua desa memiliki karakteristik yang berbeda. Sekeladi memiliki luas tanaman menghasilkan sebesar 85 hektare dengan rata-rata produktivitas 20,75 ton per hektare, sedangkan Sintong Makmur hanya memiliki luas 65 hektare tetapi menghasilkan rata-rata produktivitas sebesar 27,13 ton per hektare.

3.4 Pola Tren dan Musiman Produksi

Analisis produksi bulanan menunjukkan adanya pola musiman yang relatif konsisten selama periode 2023–2025. Produksi cenderung rendah pada awal tahun, meningkat mulai Maret, mencapai tingkat tinggi pada Mei serta Agustus–September, dan kembali menurun pada November–Desember.

Tabel 6. Pola produksi TBS berdasarkan bulan selama 2023–2025

Bulan	Total tiga tahun	Rata-rata total bulanan	Proporsi produksi	Indeks musiman
Januari	3.281,49 ton	1.093,83 ton	7,34%	88,08
Februari	3.154,82 ton	1.051,61 ton	7,06%	84,68
Maret	3.581,85 ton	1.193,95 ton	8,01%	96,14
April	3.870,58 ton	1.290,19 ton	8,66%	103,89
Mei	4.134,64 ton	1.378,21 ton	9,25%	110,98
Juni	3.710,04 ton	1.236,68 ton	8,30%	99,59
Juli	4.018,61 ton	1.339,54 ton	8,99%	107,87
Agustus	4.183,90 ton	1.394,63 ton	9,36%	112,30
September	4.181,57 ton	1.393,86 ton	9,35%	112,24
Oktober	3.944,99 ton	1.315,00 ton	8,82%	105,89
November	3.475,62 ton	1.158,54 ton	7,77%	93,29
Desember	3.167,84 ton	1.055,95 ton	7,09%	85,03

Variabel dasar yang digunakan dalam penelitian terdiri atas luas tanaman menghasilkan dan umur tanaman dalam pembentukan fitur temporal dapat dilihat pada table dibawah.

Tabel 7. Variabel Hasil Rekayasa Fitur

Fitur	Sumber	Rentang atau satuan	Fungsi
Luas tanaman menghasilkan	Data asli	8-85 ha	Merepresentasikan kapasitas lahan produktif
Umur tanaman	Data asli	9-11 tahun	Merepresentasikan karakteristik periode tanaman
MonthSin	Nomor bulan	-1 sampai 1	Merepresentasikan pola bulan secara siklis
MonthCos	Nomor bulan	-1 sampai 1	Melengkapi posisi bulan dalam siklus tahunan
Lag1	Produksi t-1	Ton	Menunjukkan produksi satu bulan sebelumnya
Lag2	Produksi t-2	Ton	Menunjukkan produksi dua bulan sebelumnya
Lag3	Produksi t-3	Ton	Menunjukkan produksi tiga bulan sebelumnya
MA3	Rata-rata produksi sampai t-3	Ton	Merepresentasikan kecenderungan produksi jangka pendek
Growth	Perubahan produksi terhadap t-2	Rasio	Menunjukkan arah dan besar perubahan produksi
SD3	Simpangan baku produksi tiga bulan	Ton	Merepresentasikan volatilitas produksi jangka pendek
Produksi TBS	Data asli	Ton	Variabel target

Hasil deskriptif fitur temporal pada 528 observasi ditampilkan pada Tabel 9.

Tabel 8. Statistik Ringkas Hasil Rekayasa Fitur

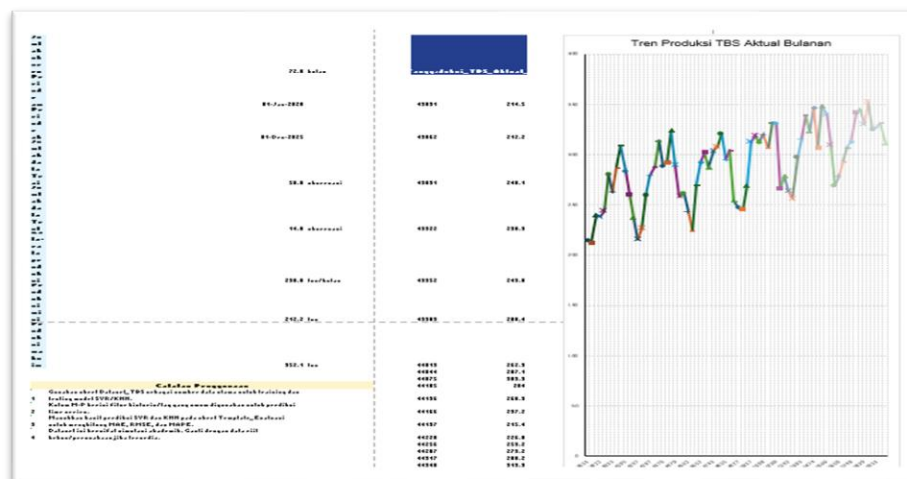
Fitur	Rata-rata	Simpangan baku	Minimum	Maksimum
Luas tanaman menghasilkan	39,63 ha	23,56 ha	8,00 ha	85,00 ha
Umur tanaman	10,09 tahun	0,79 tahun	9 tahun	11 tahun
Lag1	78,66 ton	46,35 ton	12,88 ton	173,23 ton
Lag2	78,40 ton	46,29 ton	12,31 ton	173,23 ton
Lag3	77,92 ton	46,10 ton	12,31 ton	173,23 ton
MA3	78,33 ton	45,94 ton	13,29 ton	169,26 ton
Growth	0,79%	9,03%	-19,21%	21,33%
SD3	4,17 ton	3,34 ton	0,05 ton	18,06 ton
Produksi target	78,52 ton	46,29 ton	12,85 ton	173,23 ton

Dibandingkan F0, MAE SVR berkurang sebesar 53,95%, sedangkan MAE KNNR berkurang sebesar 47,42%. Hasil tersebut menjadi bukti utama bahwa rekayasa fitur temporal berkontribusi terhadap peningkatan kinerja prediksi.

Tabel 9. Ringkasan Kontribusi Fitur Temporal Terhadap Kinerja Model

Perubahan skenario	Informasi yang ditambahkan	Dampak pada SVR		Dampak pada KNNR		Interpretasi
F0 ke F1	Kalender siklis	MAE 12,10%	turun	MAE 17,72%	turun	Pola bulan memberikan informasi musiman
F1 ke F2	Lag1	MAE 40,66%	turun	MAE 25,08%	turun	Produksi bulan sebelumnya merupakan fitur paling dominan
F2 ke F3	Lag2 dan Lag3	MAE 10,33%	turun	MAE turun 0,18%		Riwayat tambahan lebih berguna bagi SVR
F3 ke F4	MA3, Growth, dan SD3	MAE turun 1,56%		MAE 14,55%	turun	Ringkasan tren dan volatilitas lebih berguna bagi KNNR
F0 ke F4	Seluruh fitur temporal	MAE 53,95%	turun	MAE 47,42%	turun	Rekayasa fitur temporal meningkatkan kedua algoritma

KNNR lebih baik daripada SVR pada F0 dan F1. Pada F0, KNNR menghasilkan MAE 18,61% lebih rendah daripada SVR. Pada F1, selisih MAE meningkat menjadi sekitar 23,82%. Luas tanaman menghasilkan membentuk kelompok desa dengan skala produksi yang relatif berbeda. KNNR dapat menggunakan kedekatan luas lahan untuk mencari observasi dengan tingkat produksi yang serupa. Setelah MonthSin dan MonthCos ditambahkan, model juga dapat mencari observasi pada posisi musim yang berdekatan. SVR harus membentuk satu fungsi yang menjelaskan seluruh desa dan seluruh bulan secara bersamaan. Ketika jumlah fitur masih terbatas, fungsi tersebut belum memiliki informasi yang cukup untuk merepresentasikan variasi lokal pada setiap kombinasi desa dan bulan.



Gambar 2. Tren Produksi TBS

4. KESIMPULAN

Hasil analisis perbandingan Support Vector Regression (SVR) dan K-Nearest Neighbor Regression (KNN) untuk memprediksi produksi Tandan Buah Segar (TBS) kelapa sawit bulanan di CV. Bayu Sukma Jaya menunjukkan bahwa model SVR dapat menggunakan fungsi kernel (linear/RBF/polynomial) untuk menangani pola data produksi TBS yang bersifat non-linear. Tingkat akurasi prediksi SVR ditunjukkan dengan nilai error (MAE, RMSE, Pemilihan nilai k, atau jumlah tetangga ideal, sangat memengaruhi kinerja model KNN. Nilai k yang berlebihan menyebabkan overfitting, sementara nilai k yang berlebihan menyebabkan underfitting. Jarak Euclidean/Manhattan juga memengaruhi hasil prediksi.

Untuk memprediksi produksi tandan buah segar bulanan pada CV. Bayu Sukma Jaya, algoritma Support Vector Regression dan K-Nearest Neighbor Regression dapat digunakan. Meskipun masing-masing model memiliki kemampuan untuk mempelajari pola produksi dari data sebelumnya, mereka tidak seakurat satu sama lain. Hasil evaluasi dengan MAE, RMSE, MAPE, dan R2 menunjukkan bahwa model Support Vector Regression (SVR) memiliki koefisien determinasi yang lebih tinggi dan nilai kesalahan yang lebih rendah daripada model K-Nearest Neighbor Regression (KNN).

REFERENCES

- [1] K. Badan and P. Statistik, "No Title," pp. 1–564, 2025.
- [2] A. Widiarni, "Penerapan Algoritma Support Vector Regression dalam Memprediksi Produksi dan Produktivitas Kelapa Sawit," vol. 7, no. April, pp. 864–872, 2023, doi: 10.30865/mib.v7i2.6089.
- [3] A. Al Mahmud, Z. Hossan, A. Tiwari, and R. Khatoun, "Reviewing the Integration of RFID and IoT in Supply Chain Management : Enhancing Efficiency and Visibility," vol. 3576, pp. 409–437, 2025.
- [4] K. Neighbors, "Penerapan Data Mining Untuk Klasifikasi Penduduk Miskin Di Kabupaten Labuhanbatu Menggunakan Random Forest Dan," vol. 6, no. 2, pp. 23–36, 2025, doi: 10.47065/bit.v5i2.1783.
- [5] I. Maulana and A. Efendi, "Virtual – Augmented Reality (VAR) for Science Learning : Development and Impact on Students ' HOTS Skills," vol. 23, no. 1, pp. 129–142, 2025.
- [6] P. H. Putra, M. S. Novelan, and M. Rizki, "ANALYSIS K-NEAREST NEIGHBOR METHOD IN CLASSIFICATION OF VEGETABLE QUALITY BASED ON COLOR," vol. 3, no. 2, pp. 126–132, 2022.
- [7] Z. Sitorus, A. Marsya, D. Ramatika, and R. S. Siburian, "Analysis of Online Shopping Addiction Level Using the K-nearest Neighbor Algorithm at SMK Negeri 1 Tanjung Pura," *J. Inf. Technol. Comput. Sci. Electr. Eng.*, vol. 1, no. 2, pp. 134–138, 2024, doi: 10.61306/jitcse.v1i2.
- [8] F. Haplosapristis, "GAMBUT DI LABUHAN BATU SOIL FERTILITY AND PHOTOSYNTHESIS RATE OF OIL PALM SHOWING," vol. 27, no. 2, pp. 127–140, 2019.
- [9] D. S. Prasvita, M. M. Santoni, R. Wirawan, and N. Trihastuti, "Klasifikasi pohon kelapa sawit pada data fusi citra lidar dan foto udara menggunakan convolutional neural network," vol. 06, pp. 406–415, 2021.
- [10] A. R. M. Akbar, A. D. Wibowo, and B. H. Asyikin, "Risk Assessment for Long - Term Injury of Oil Palm Harvester," vol. 14, no. 2, pp. 594–603, 2025.
- [11] R. Hasibuan, F. Roosmawati, and T. Z. Aznur, "Analisis Harga Pokok Produksi Tandan Buah Segar (TBS) pada Kebun Pengarungan PT . Asam Jawa Kabupaten Labuhanbatu Selatan," vol. 5, no. April, 2026.
- [12] N. W. Wardani, P. Yayasan, and K. Menulis, *No Title*.
- [13] Z. Sitorus, E. Hariyanto, and F. Kurniawan, "Implementasi Machine Learning Pada Sistem Pemetaan Daerah Rawan Banjir Di Desa Pahlawan Kabupaten Batu Bara," vol. 3, no. 3, pp. 285–290, 2022.
- [14] J. M. Polgan, A. Putera, U. Siahaan, M. Iqbal, and M. Syahputri, "Classification Of Pistachio Varieties Using Machine Learning Algorithms," vol. 14, pp. 1452–1457, 2025.
- [15] B. Djehiche and B. Löfdahl, "Quantum Support Vector Regression for Disability Insurance," 2021.
- [16] H. Hissou, S. Benkirane, A. Guezaz, and M. Azrour, "A Novel Machine Learning Approach for Solar Radiation Estimation," 2023.