



# Optimalisasi Ketahanan Pangan: Perbandingan Metode *Machine Learning* dan *Time Series* dalam Memprediksi Produksi Padi di Jawa Tengah

Ni Komang Ayu Mita<sup>1</sup>, Muhammad Fajar Siddiq<sup>2</sup>, Anastasia Laurnt<sup>3</sup>, Rissa Erviana<sup>4</sup>,  
Robert Kurniawan<sup>5</sup>

<sup>1, 2, 3, 4</sup>Prodi Statistika, Politeknik Statistika STIS, Jakarta, Indonesia

<sup>5</sup>Prodi Komputasi Statistika, Politeknik Statistika STIS, Jakarta, Indonesia

<sup>1</sup>[212112252@stis.ac.id](mailto:212112252@stis.ac.id)

<sup>2</sup>[212112211@stis.ac.id](mailto:212112211@stis.ac.id)

<sup>3</sup>[212111887@stis.ac.id](mailto:212111887@stis.ac.id)

<sup>4</sup>[212112328@stis.ac.id](mailto:212112328@stis.ac.id)

<sup>5</sup>[robertk@stis.ac.id](mailto:robertk@stis.ac.id)

Corresponding author email: [212112252@stis.ac.id](mailto:212112252@stis.ac.id)

**Abstract:** Food security remains a challenge for Indonesia. Despite the Indonesian government's prioritization of long-term food security policies, issues of malnutrition and food insecurity persist. Rice serves as the primary food source for over 50% of the global population and significantly contributes to global food security. However, the impact of climate change on rice yields and the accuracy of traditional regression methods in predicting rice yields are still questioned. This study aims to compare machine learning methods (Support Vector Regression and Random Forest) with time series methods (SARIMAX) using rice production as the response variable and considering predictor variables such as land area, NDWI, temperature, and rainfall. The study focuses on Central Java, one of the provinces with the highest rice production in Indonesia. The Random Forest method provided the best results with the highest  $R^2$  value and the lowest MAPE at 97.6% and 7.827, respectively.

**Keywords:** rice production, SVR, RFR, SARIMAX, food security

**Abstrak:** Ketahanan pangan masih menjadi tantangan bagi Indonesia. Meskipun pemerintah Indonesia telah menjadikan ketahanan pangan sebagai prioritas kebijakan jangka panjang, kekurangan gizi dan kerawanan pangan tetap menjadi permasalahan. Padi menjadi sumber pangan utama bagi lebih dari 50% populasi dunia dan signifikan berkontribusi terhadap ketahanan pangan dunia. Namun, dampak perubahan iklim terhadap hasil panen padi dan keakuratan metode regresi tradisional dalam memprediksi hasil padi masih dipertanyakan. Penelitian ini bertujuan untuk membandingkan metode machine learning (Support Vector Regression dan Random Forest) dan metode time series (SARIMAX) dengan variabel respons produksi padi serta mempertimbangkan variabel prediktor luas lahan, NDWI, suhu, dan curah hujan dengan lokus Jawa Tengah sebagai salah satu provinsi dengan produksi padi tertinggi di Indonesia. Metode Random Forest memberikan hasil terbaik dengan nilai  $R^2$  tertinggi dan MAPE terkecil sebesar 97,6% dan 7,827.

**Kata kunci:** produksi padi, SVR, RFR, SARIMAX, ketahanan pangan.

## I. PENDAHULUAN

Ketahanan pangan masih menjadi tantangan bagi negara-negara berpendapatan rendah dan menengah seperti Indonesia. Pemerintah Indonesia menjadikan ketahanan pangan sebagai prioritas kebijakan jangka panjang. Namun, kekurangan gizi dan kerawanan pangan masih menjadi masalah di Indonesia, serta kebijakan ketahanan pangan sangatlah penting untuk dilanjutkan [1]. SDG-2 bertujuan untuk "Mengakhiri kelaparan, mencapai ketahanan pangan dan peningkatan nutrisi serta mendorong pertanian berkelanjutan". Meskipun negara-negara miskin cenderung sangat bergantung pada kegiatan pertanian, produksi dan konsumsi pangan merupakan basis perekonomian dan memengaruhi kehidupan bermasyarakat [2]. Dalam bidang pertanian, padi merupakan sumber pangan utama bagi lebih dari 50% populasi dunia dan memberikan kontribusi signifikan terhadap ketahanan pangan bagi populasi dunia [3]. Indonesia merupakan negara yang menjadikan sektor pertanian sebagai sektor utama dalam pembangunan ekonomi. Namun, dampak perubahan iklim terhadap hasil panen padi masih belum jelas



dan belum tereksplorasi, dan prediksi hasil padi dengan menggunakan metode regresi tradisional masih dipertanyakan [4]. Oleh karena itu, untuk mengatasi tantangan perubahan iklim global dan menjamin ketahanan pangan, penting untuk menyelidiki variabel iklim terhadap hasil produksi padi dan meningkatkan keakuratan prediksi padi.

Pendekatan hasil estimasi panen dalam beberapa tahun terakhir masih menggunakan teknik tradisional, termasuk model simulasi tanaman berorientasi proses, dan model berbasis statistik antara hasil tanaman dan faktor penjelas [5]. Metode tradisional berbasis statistik dapat memberikan alternatif untuk prediksi hasil karena perhitungannya yang lebih sederhana dan kekuatan penjelasan yang tinggi [6]. Namun, masih banyak masalah dengan model regresi empiris tradisional karena generalisasi spasialnya yang biasanya terlokalisasi dan terbatas [7]. Oleh karena itu, terdapat kebutuhan untuk mengembangkan pendekatan baru untuk memprediksi hasil panen yang andal, tepat waktu, dan hemat biaya di wilayah yang luas. Dibandingkan dengan model statistik tradisional, *Machine Learning (ML)* memiliki kemampuan lebih baik untuk menangani pola hubungan yang kompleks antara variabel prediktor dan respon. Dalam beberapa tahun terakhir, metode ini semakin banyak digunakan dalam bidang penelitian pertanian (klasifikasi tanaman, pemantauan pertumbuhan, prediksi hasil panen) di berbagai negara [8]. Misalnya, memprediksi hasil panen padi menggunakan *Least Absolute Shrinkage and Selection Operator (LASSO) regression*, *machine learning (Random Forest, RF)*, dan *deep learning (Long Short - Term Memory Networks, LSTM)* di tingkat kabupaten di seluruh Tiongkok [9]. Contoh lain yaitu membandingkan tiga metode ML termasuk *Support Vector Machine (SVM)*, *random forest (RF)* dan *Neural Network (NN)*, serta metode regresi tradisional *Least Absolute Shrinkage and Selection Operator-LASSO* untuk prediksi hasil gandum di Australia [5].

Meskipun telah banyak penelitian yang dilakukan di bidang prediksi produksi padi, sebagian besar dari penelitian tersebut lebih banyak membandingkan beberapa metode *Machine Learning* dan memprediksi menggunakan metode tradisional seperti *Time Series* tanpa memperhitungkan variabel prediktor yang mungkin mempengaruhi hasil prediksi. Namun, masih sedikit penelitian yang secara langsung membandingkan efektivitas metode *Machine Learning* dengan metode *Time Series* dalam memprediksi produksi padi. Oleh karena itu, sejauh mana metode *Machine Learning* dapat mengungguli metode tradisional seperti *Time Series* dalam memprediksi produksi padi dengan mempertimbangkan variabel prediktor yang dilakukan dengan cara membandingkan hasil evaluasi dari model yang terbentuk.

Tujuan dari penelitian ini adalah untuk mengevaluasi dan membandingkan hasil prediksi produksi padi menggunakan pendekatan *Machine Learning* yaitu *Support Vector Regression (SVR)* dan *Random Forest* dengan *Time Series* yaitu SARIMAX. Beberapa indikator yang digunakan dalam prediksi produksi padi yaitu luas panen, *Normalized Difference Water Index (NDWI)*, suhu permukaan tanah, dan curah hujan secara bulanan. Jawa Tengah dipilih sebagai lokasi penelitian karena termasuk provinsi dengan produksi padi terbesar di Indonesia, mencapai 9,356 juta ton pada tahun 2022 [10]. Namun, jumlah produksi tersebut menurun dari tahun 2018 yang mencapai 10,499 juta ton [10]. Maka dari itu, prediksi padi yang tepat sangat penting untuk pengambilan keputusan dan kebijakan terkait ketahanan pangan dan pertanian berkelanjutan dalam wilayah regional Jawa Tengah.

## II. METODE PENELITIAN

Jawa Tengah adalah salah satu pusat produksi padi terbesar di Indonesia, sehingga analisis produksi padi di wilayah ini sangat penting. *Support Vector Regression (SVR)* dan *Random Forest (RFR)* adalah beberapa metode *machine learning* yang dapat digunakan untuk memahami dan memprediksi produksi



padi, sedangkan ARIMAX adalah salah satu metode *time series* yang dapat digunakan untuk memprediksi produksi padi dengan tambahan variabel prediktor. Bab ini akan membahas landasan teori untuk ketiga metode tersebut beserta definisi, prinsip dasar, dan aplikasinya. Ini juga akan membahas variabel bebas yang digunakan dalam analisis: luas panen padi, indeks perbedaan normalisasi air (NDWI), curah hujan, dan suhu tanah.

### II.1. *Produksi Padi*

Produksi padi adalah jumlah padi yang dihasilkan oleh lahan pertanian dalam satu periode tanam. Ini adalah indikator utama kinerja sektor pertanian padi dan berkontribusi besar terhadap ketahanan pangan, terutama di negara seperti Indonesia yang mengonsumsi beras secara signifikan. Pada penelitian ini, produksi padi berperan sebagai variabel respons. Data produksi padi diambil dari publikasi Badan Pusat Statistik (BPS) yaitu data bulanan tahun 2018-2022 di Provinsi Jawa Tengah, sehingga data yang digunakan adalah sebanyak 60 *series*.

### II.2. *Luas Panen Padi*

Luas panen padi mengacu pada total area yang menghasilkan padi. Penurunan luas panen yang difungsikan untuk pertanian padi berkorelasi positif dengan penurunan produksi padi, terutama di daerah yang perubahan penggunaan lahan cepat terjadi [11]. Data ini diambil dari publikasi Badan Pusat Statistik (BPS), yaitu data bulanan tahun 2018-2022 di Provinsi Jawa Tengah.

### II.3. *Normalized Difference Water Index (NDWI)*

NDWI adalah indeks untuk menghitung jumlah air yang ada dalam vegetasi dengan menggunakan data citra satelit. NDWI dihitung menggunakan rasio spektral dari cahaya hijau dan cahaya inframerah (NIR) yang dipantulkan oleh tanaman. Jumlah air dapat memengaruhi produksi padi, NDWI yang tinggi menunjukkan bahwa tanaman menerima air yang cukup, yang mendukung pertumbuhan dan hasil yang optimal [12]. Data NDWI diambil dari salah satu produk citra satelit yang dikelola oleh *United States Geological Survey* (USGS), yaitu LANDSAT/LC08/C02/T1\_TOA.

### II.4. *Suhu*

Suhu tanah adalah suhu pada permukaan tanah atau pada kedalaman tertentu yang memengaruhi berbagai proses biologis dan fisik di dalam tanah, seperti penyerapan air dan nutrisi oleh akar tanaman. Suhu dapat memengaruhi produksi padi, suhu yang meningkat 1–2 °C selama fase pengisian biji padi dapat menurunkan produksi padi hingga 10–20 persen [13]. Data diambil dari dataset iklim global yang dikelola oleh proyek *Experimental Program to Stimulate Competitive Research* (EPSCoR) yaitu IDAHO\_EPSCOR/TERRACLIMATE.

### II.5. *Curah Hujan*

Salah satu variabel iklim yang paling memengaruhi pertanian, terutama untuk tanaman padi adalah curah hujan, yang merupakan jumlah air yang jatuh ke permukaan bumi dalam bentuk hujan. Curah hujan yang memadai dan terdistribusi dengan baik akan meningkatkan produktivitas padi, sementara curah hujan yang terlalu sedikit atau tidak teratur dapat mengakibatkan penurunan hasil [14]. Data curah hujan diambil dari *Climate Hazards Group InfraRed Precipitation with Station Data* (CHIRPS) yang memiliki resolusi tinggi dan cakupan temporal yang panjang.

### II.6. *Optimasi Hyperparameter*

Parameter *hyperparameter* diatur nilainya sebelum proses pelatihan model *machine learning*. Parameter ini berbeda dari parameter model yang diperoleh selama pelatihan. *Hyperparameter* memengaruhi kinerja model dan proses pelatihan. Jumlah pohon dalam *Random Forest* dan nilai C



dalam *Support Vector Machine* adalah dua contoh *hyperparameter*. *Hyperparameter* penting dalam SVR meliputi *Cost (C)*, *Gamma ( $\gamma$ )*, *Epsilon ( $\epsilon$ )*, dan *Kernel*. *Cost* adalah parameter yang berfungsi menentukan seberapa banyak *error* yang diizinkan dalam model. Nilai *C* yang tinggi berusaha untuk mengklasifikasikan semua poin pelatihan secara benar, yang mengurangi margin error, sedangkan nilai *C* yang rendah memungkinkan margin yang lebih besar dengan konsekuensi yang lebih sedikit untuk misklasifikasi [15]. *Gamma* adalah parameter yang menentukan seberapa jauh pengaruh dari satu data poin. *Gamma* yang lebih tinggi menunjukkan bahwa radius pengaruh dari poin data yang dipilih sebagai vektor pendukung lebih dekat dan sebaliknya [15]. *Epsilon* adalah *hyperparameter* yang mengontrol batas kesalahan prediksi nilai  $\epsilon$  yang lebih tinggi menunjukkan bahwa model SVR lebih tahan terhadap kesalahan prediksi [15]. Fungsi yang mengubah data non-linear ke dalam ruang yang lebih besar di mana data dapat dipisahkan secara linear dikenal sebagai kernel. Jenis kernel yang digunakan menentukan fungsi transformasi yang digunakan [15].

Sedangkan dalam RFR, *hyperparameter* berupa *Mytry (Number of Predictor Variables to Sample at Each Split)*, *Splitrule (Splitting Rule)*, *Min.Node.Size (Minimum Size of Terminal Nodes)*, *Ntree (Number of Trees)*. *Mytry* adalah *hyperparameter* yang mengatur banyaknya variabel prediktor yang akan dipilih secara acak untuk setiap pembagian. Jumlah metrik yang lebih besar menunjukkan bahwa model RFR lebih rumit dan memiliki kemampuan untuk menangani masalah data non-linier. *Hyperparameter splitrule* mengontrol aturan pembagian. Nilai *splitrule* yang lebih besar menunjukkan bahwa model RFR lebih kompleks dan dapat menangani masalah data non-linier. *Hyperparameter min.node.size* mengatur ukuran minimum node terminal. Nilai *min.node.size* yang lebih tinggi menunjukkan bahwa model RFR lebih kompleks dan dapat menangani masalah data non-linier. Jumlah pohon dalam model RFR diatur oleh *hyperparameter* pohon. Nilai pohon yang lebih besar menunjukkan bahwa model RFR lebih kompleks dan dapat menangani masalah data non-linier [16].

## II.7. SVR

*Support Vector Regression (SVR)* adalah teknik regresi yang berasal dari *Support Vector Machine (SVM)*. SVM pertama kali dibuat untuk tugas klasifikasi, tetapi SVR mengubahnya untuk menangani tugas regresi. Salah satu keuntungan utama SVR adalah dapat menangani hubungan non-linear menggunakan kernel. Selain itu, ia memiliki kemampuan generalisasi yang sangat baik, dengan akurasi prediksi yang tinggi [17]. Tujuan SVR adalah untuk mengurangi kompleksitas model dengan menemukan fungsi yang memiliki deviasi maksimal  $\epsilon$  dari nilai sebenarnya untuk semua data *training*. Metode ini menggunakan prinsip margin, yang juga digunakan oleh SVM, tetapi digunakan dalam regresi. Prinsip utama SVR adalah menggunakan fungsi kernel untuk memetakan data ke ruang yang lebih besar, kemudian menemukan *hyperplane* terbaik yang memprediksi nilai target. Persamaan fungsi regresi untuk SVR dapat dirumuskan sebagai berikut:

$$f(x) = w^T \varphi(x) + b \quad (1)$$

Dimana  $\varphi(x)$  adalah fungsi pemetaan dari  $x$  pada dimensi yang lebih tinggi,  $w$  adalah vektor pembobot, dan  $b$  adalah bias.

Untuk menangani data non-linear, SVR menggunakan fungsi kernel, yang menghubungkan data asli ke ruang fitur berdimensi lebih besar. Beberapa fungsi kernel yang paling umum digunakan adalah *linear kernel*, *polynomial kernel*, dan *radial basis function kernel*. Penelitian ini menggunakan *radial basis function kernel* dengan persamaan:

$$K(x_i, x_j) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (2)$$

Keterangan:

$x$  dan  $x_i$  = data

$\sigma$  = konstanta dimensi

### II.8. RFR

*Random Forest* adalah metode pembelajaran kelompok yang menggabungkan hasil dari banyak pohon keputusan yang dibangun dari subset data yang dipilih secara acak untuk tugas klasifikasi dan regresi. Prinsip dasar dari *Random Forest* adalah bahwa menggabungkan hasil dari banyak model pohon keputusan yang agak sederhana akan meningkatkan prediksi dan mengurangi *overfitting*. Algoritma *Random Forest* memiliki beberapa langkah utama:

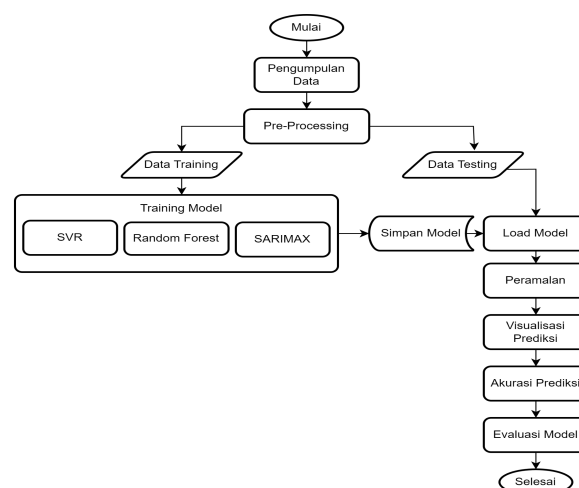
1. Pembentukan subset data (*bootstrap sampling*): Dengan menggunakan teknik *bootstrap*, algoritma membuat beberapa subset data dari dataset asli; setiap subset dipilih secara acak dan diganti dengan data asli.
2. Pembentukan pohon keputusan: Sebuah pohon keputusan dibangun pada setiap subset data dan setiap pohon tumbuh tanpa pemangkasan sampai mencapai ukurannya yang penuh. Setiap jarak di simpul pohon dipilih dari subset fitur acak yang diberikan.
3. Prediksi *ensemble*: Untuk regresi, hasilnya adalah rata-rata dari keseluruhan prediksi pohon.

### II.9. ARIMAX

ARIMAX (*AutoRegressive Integrated Moving Average with Exogenous Inputs*) adalah lanjutan dari model ARIMA yang memasukkan variabel eksogen (atau input tambahan) ke dalam proses prediksi. Model ini menggabungkan pendekatan ARIMA dengan variabel input eksternal untuk membuat prediksi yang lebih akurat. ARIMAX digunakan untuk menganalisis data pertanian karena efektif dalam menangkap pola musiman dan tren dalam data.

### II.10. Alur Analisis

Proses prediksi produksi padi pada penelitian ini digambarkan melalui diagram alur yang dapat dilihat pada gambar 1.



**Gambar 1.** Diagram Alur Prediksi Produksi Padi

Proses pengerjaan diawali dengan mengumpulkan data yang digunakan sebagai variabel respons dan variabel prediktor. Kemudian, dilanjutkan tahapan *pre-processing* baik untuk metode *machine learning* (SVR dan RFR) maupun metode *time series* (SARIMAX). *Dataset* yang digunakan merupakan gabungan dari variabel respons, variabel prediktor, dan *series* waktu (bulanan) dan sudah berupa *clean data*. Oleh karena itu, tidak diperlukan tahapan *cleaning data*, seperti pengisian data yang hilang,





transformasi data, normalisasi data, maupun mengubah tipe data. Kemudian, dilakukan penetapan *seed* acak. Fungsi dari proses ini adalah reproduksibilitas dan konsistensi pengujian. Dengan penetapan tersebut, pembagian data *training* dan data *testing* secara acak dapat diulang dengan cara yang sama sehingga memberikan hasil yang konsisten.

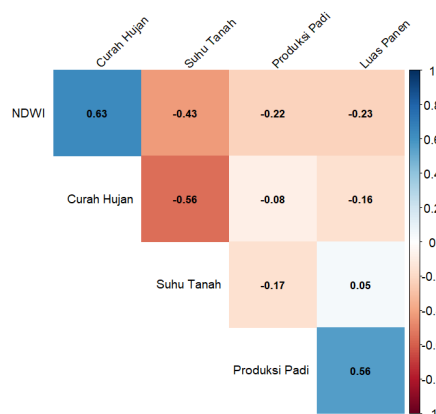
Tahapan selanjutnya dalam *pre-processing* adalah membagi data menjadi data *training* dan data *testing* yang diterapkan pada metode *machine learning* maupun metode *time series*. Pada metode *machine learning*, tujuan dari pembagian data ini adalah untuk memastikan bahwa model yang dibentuk dapat menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya. Sementara pada metode *time series*, data *testing* nantinya akan dibandingkan dengan hasil prediksi menggunakan model SARIMAX, sehingga diperoleh nilai evaluasi model yang bisa dibandingkan dengan metode *machine learning*. Penelitian ini menggunakan data *training* dan data *testing* dengan perbandingan 60:40. Penggunaan perbandingan tersebut didasari atas keterbatasan jumlah data untuk pemodelan *time series*. Jumlah data *testing* dengan perbandingan tersebut adalah sebanyak 24 *series*. Dengan demikian, masih dapat dilakukan pemodelan menggunakan *seasonal* dengan periode 12 apabila terdeteksi terdapat pola *seasonal*.

Tahapan selanjutnya adalah melakukan pemodelan SVR, RFR, dan SARIMAX menggunakan data *training*. Pada model SARIMAX, dilakukan deteksi pola *seasonal* terlebih dahulu melalui grafik produksi padi tahun 2018-2022 dan dilanjutkan dengan menentukan ordo yang akan digunakan melalui grafik ACF dan PACF. Dari pemodelan data *training*, kemudian dilakukan pemodelan dengan model yang sama tetapi menggunakan data *testing* untuk ketiga model. Dilanjutkan dengan melakukan prediksi untuk dibandingkan dengan data *testing*, sehingga dapat diperoleh visualisasi dari perbandingan tersebut. Akurasi prediksi juga didapatkan melalui nilai  $R^2$  dan MAPE. Tahapan akhir, evaluasi dilakukan dalam perbandingan metode *machine learning* dan *time series*.

### III. HASIL DAN PEMBAHASAN

#### III.1. Hasil

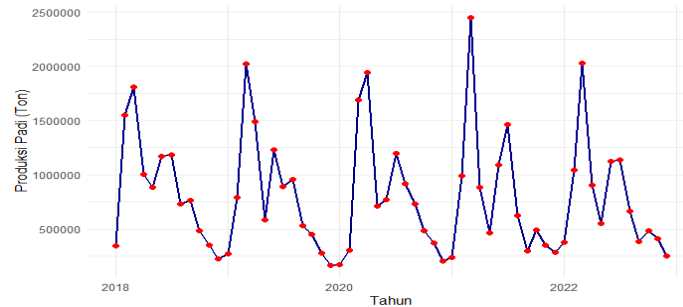
##### III.1.1. Analisis Deskriptif



Gambar 2. Heat Map Korelasi

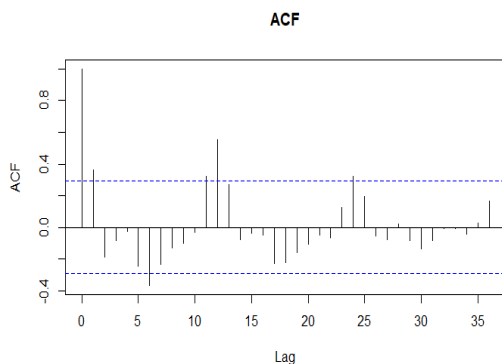
Kekuatan dan arah hubungan antar variabel dapat dilihat dari *heat map* korelasi pada gambar 2. Produksi padi sebagai variabel respons memiliki hubungan positif dengan luas panen. Artinya, meningkatnya luas panen padi diiringi dengan peningkatan produksi padi. Sementara itu, NDWI, curah hujan, dan suhu tanah memiliki hubungan yang negatif dan relatif lemah dengan

produksi padi. Dengan artian meningkatnya NDWI (jumlah air dalam vegetasi), curah hujan, dan suhu tanah diiringi dengan penurunan produksi padi.

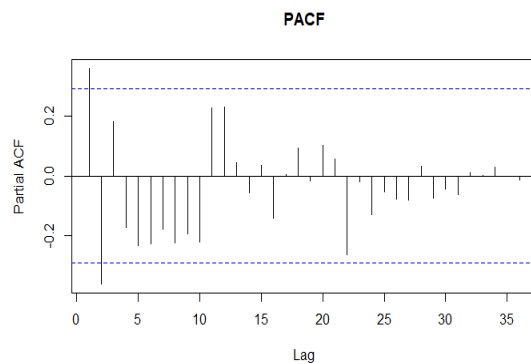


**Gambar 3.** Produksi Padi Jawa Tengah 2018-2022

Berdasarkan grafik pada gambar 3, produksi padi di Jawa Tengah tahun 2018 hingga 2022 secara keseluruhan terlihat fluktuatif. Jika diamati lebih lanjut, secara umum produksi padi di Jawa Tengah menunjukkan pola musiman. Produksi padi selalu meningkat hingga mencapai produksi tertinggi per tahun pada bulan Maret. Oleh karena itu, dapat dikatakan bahwa data produksi padi memiliki pola musiman dengan periode 12 bulan. Hal ini kemudian digunakan pada pemodelan ARIMAX sebagai komponen *seasonal*, sehingga model yang dibentuk menjadi  $SARIMAX(p, d, q)(P, D, Q)^s$  dengan  $s$  sama dengan 12.



**Gambar 4a.**



**Gambar 4b.**

**Gambar 2.** Plot ACF dan PACF Produksi Padi (4a) Plot ACF (4b) Plot PACF

Gambar 4a dan 4b digunakan untuk menentukan ordo model ARIMAX. Pada grafik ACF, terlihat *cut off* pada lag ke-1 atau lag ke-2 yang berarti ordo dari *moving average* ( $q$ ) adalah 1 atau 2. Sementara itu, grafik PACF menunjukkan *cut off* pada lag ke-2 yang berarti ordo dari *autoregressive* ( $p$ ) adalah 2. Pola musiman terlihat dari grafik PACF dengan *cut off* pada lag ke-12, yang berarti ordo dari *autoregressive seasonal* adalah 12. Dengan demikian, akan dibandingkan model  $SARIMAX(2,0,1)(1,0,0)^{12}$  dan  $SARIMAX(2,0,2)(1,0,0)^{12}$ .

### III.1.2. Optimasi Hyperparameter Menggunakan GridsearchCV

Penelitian ini melakukan optimasi *hyperparameter* dengan metode *gridSearchCV*. Metode ini menekankan pada pencarian parameter secara menyeluruh dari nilai *hyperparameter* yang telah ditentukan di awal (validasi silang). Optimasi *hyperparameter* untuk model SVR menggunakan beberapa parameter, seperti *cost* ( $C$ ), *gamma* ( $\gamma$ ), *epsilon* ( $\epsilon$ ), dan *kernel*. Dalam hal ini, *cost* ( $C$ ) berisi 0.1, 1, dan 10. *Gamma* ( $\gamma$ ) berisi nilai 0.01, 0.1, dan 1. *Epsilon* ( $\epsilon$ ) berisi nilai 0, 0.1, dan 1.



Beberapa *kernel* yang digunakan yaitu linear, radial, dan polynomial. Validasi silang menggunakan 10 lipatan, sehingga menghasilkan hasil sebagai berikut.

**Tabel 1.** Hasil Optimasi *Hyperparameter* SVR

<i>Hyperparameter</i>	Parameter Terbaik
<i>cost</i> (C)	10
<i>gamma</i> ( $\gamma$ )	0.1
<i>epsilon</i> ( $\epsilon$ )	0
<i>kernel</i>	radial

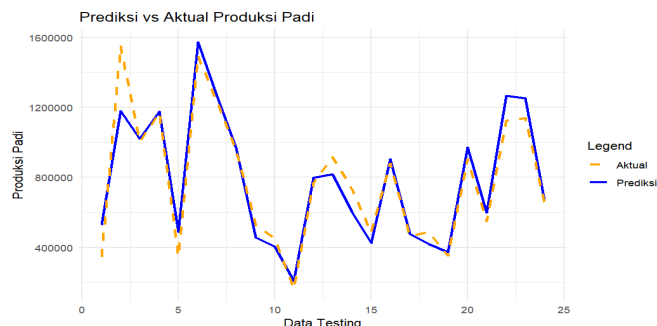
Sementara itu, beberapa parameter yang digunakan untuk model RFR, antara lain *mtry*, *splitrule*, dan *min.node.size*. Dalam hal ini, *mtry* berisi nilai 2, 3, dan 4. *Splitrule* antara *variance* dan *extratrees*. *Min.node.size* berisi nilai 1, 3, 5, dan 7. *Ntree* berisi nilai 500, 1000, dan 1500. Validasi silang menggunakan 10 lipatan, sehingga menghasilkan hasil sebagai berikut:

**Tabel 2.** Hasil Optimasi *Hyperparameter* RFR

<i>Hyperparameter</i>	Parameter Terbaik
<i>mtry</i>	4
<i>splitrule</i>	<i>variance</i>
<i>min.node.size</i>	1
<i>ntrree</i>	500

### III.1.3. Support Vector Regression

Dengan menggunakan *hyperparameter* terbaik untuk membentuk model SVR, yaitu *cost* (C) 10, *gamma* ( $\gamma$ ) 0.1, *epsilon* ( $\epsilon$ ) 0, dan kernel radial, diperoleh hasil seperti dibawah ini.



**Gambar 5.** Hasil Prediksi Produksi Padi Model *Support Vector Regression*

Gambar 5 menunjukkan hasil prediksi belum sama dengan data aktual. Dapat dilihat masih terdapat perbedaan dengan *data testing* pada bagian awal, tengah, dan akhir prediksi. Hal ini menunjukkan bahwa dalam memprediksi produksi padi menggunakan model SVR belum memberikan akurasi yang cukup baik.

### III.1.4. Random Forest Regression

Dengan menggunakan *hyperparameter* terbaik pada tabel 2, menghasilkan persentase *var explained* sebesar 82,18% yang berarti model mampu menjelaskan variasi dalam data sebesar 82,18% dari total varians. Hasil berikut dapat menjelaskan variabel yang paling berpengaruh dalam memprediksi produksi padi.

**Tabel 3.** Hasil *Random Forest Regression*

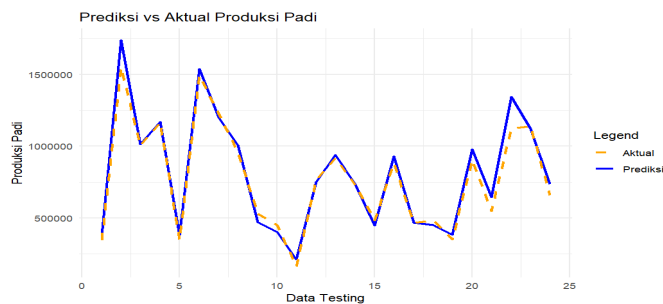
Variabel	%IncMSE	IncNodePurity
Luas Panen	51,109	$1,16.10^{13}$
NDWI	3,259	$3,06.10^{11}$





Variabel	%IncMSE	IncNodePurity
Suhu Tanah	5,402	$3,56.10^{11}$
Curah Hujan	7,398	$8,73.10^{11}$

Variabel dengan nilai %IncMSE tertinggi menunjukkan yang paling berpengaruh dalam memprediksi produksi padi dengan model RFR. Dalam konteks ini, luas panen memiliki pengaruh yang paling besar yaitu sebesar 51,109%, kemudian diikuti oleh curah hujan sebesar 7,398%. Sementara itu, nilai IncNodePurity paling tinggi menunjukkan variabel yang paling berpengaruh dalam pembentukan struktur pohon keputusan. Dapat diketahui bahwa luas panen padi juga memiliki nilai IncNodePurity paling tinggi yaitu sebesar  $1,16.10^{13}$ . Hal ini berarti luas panen memiliki pengaruh paling besar dalam pembentukan struktur pohon keputusan, kemudian diikuti juga oleh curah hujan.



**Gambar 6.** Hasil Prediksi Produksi Padi Model *Random Forest Regression*

Berdasarkan gambar 6, dapat dilihat bahwa hasil prediksi hampir sama dengan data aktual. Hal ini menunjukkan bahwa dalam memprediksi produksi padi menggunakan model RFR dapat memberikan akurasi yang baik, meskipun pada bagian akhir prediksi terdapat sedikit perbedaan dengan *data testing*.

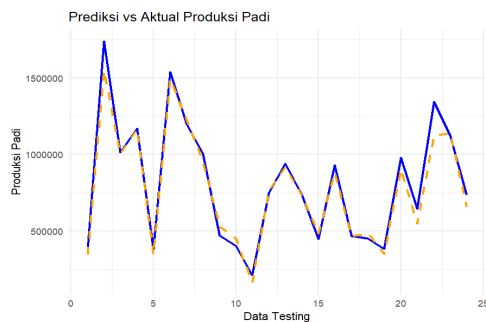
### III.1.5. SARIMAX

**Tabel 4.** Hasil Uji Stasioneritas Dengan ADF

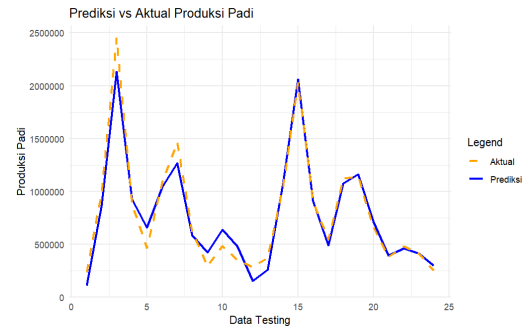
Variabel	<i>P-Value</i>
Produksi Padi	0,019*
Luas Panen	0,017*
NDWI	0,010*
Suhu Tanah	0,046*
Curah Hujan	0,010*

\*signifikan pada taraf 5%

Hasil uji ADF pada tabel 4 menunjukkan nilai *p-value* dari kelima variabel kurang dari taraf uji 5% atau 0,05, sehingga memberikan keputusan tolak hipotesis nol (data stasioner). Dengan taraf uji 5%, cukup untuk menunjukkan bahwa kelima variabel yang digunakan telah stasioner pada level. Oleh karenanya, tidak perlu dilakukan *differencing* (ordo d dan D sama dengan 0).



Gambar 7a.



Gambar 7b.

Gambar 7. Model SARIMAX (7a)  $SARIMAX(2,0,2)(1,0,0)^{I2}$  (7b)  $SARIMAX(2,0,1)(1,0,0)^{I2}$

Gambar 7a menunjukkan bahwa model  $SARIMAX(2,0,2)(1,0,0)^{I2}$  menghasilkan prediksi yang hampir sama dengan data aktualnya. Namun, terdapat sedikit perbedaan pada bagian akhir, yang mana hasil prediksi melebihi data aktual. Walaupun demikian, hasil prediksi menggunakan model  $SARIMAX(2,0,2)(1,0,0)^{I2}$  memberikan hasil akurasi yang baik. Sementara itu, gambar 7b menunjukkan bahwa model  $SARIMAX(2,0,1)(1,0,0)^{I2}$  menghasilkan prediksi yang hampir sama dengan data aktualnya. Namun, terdapat sedikit perbedaan pada bagian awal prediksi, yang mana hasil prediksi ada yang lebih dan kurang dari data aktual. Walaupun demikian, hasil prediksi menggunakan model  $SARIMAX(2,0,1)(1,0,0)^{I2}$  memberikan hasil akurasi yang baik.

Selain menggunakan grafik prediksi vs aktual, penting untuk membandingkan model SARIMAX menggunakan  $R^2$  dan MAPE. Hal ini dilakukan agar mendapatkan hasil prediksi yang baik atau mendekati dengan data aktualnya sehingga model terbaik tersebut dapat digunakan untuk memprediksi produksi padi Jawa Tengah di masa mendatang. Tabel 5 menampilkan perbandingan model  $SARIMAX(2,0,2)(1,0,0)^{I2}$  dan  $SARIMAX(2,0,1)(1,0,0)^{I2}$ . Hasil menunjukkan bahwa model  $SARIMAX(2,0,1)(1,0,0)^{I2}$  menghasilkan  $R^2$  paling tinggi dan MAPE paling kecil.

Tabel 5. Perbandingan Model SARIMAX

Model	$R^2$	MAPE (%)
$SARIMAX(2,0,2)(1,0,0)^{I2}$	0,930	21,638
$SARIMAX(2,0,1)(1,0,0)^{I2}$	0,962	16,398

### III.1.6. Hasil Evaluasi Model

Setelah memperoleh model terbaik dari SARIMAX, penting untuk membandingkan dengan model SVR dan RFR. Nilai  $R^2$  dan MAPE (*Mean Absolute Percentage Error*) digunakan untuk mengevaluasi model SVR, RFR, dan  $SARIMAX(2,0,1)(1,0,0)^{I2}$  dari 40% data *testing*. Hal ini karena  $R^2$  atau koefisien determinasi merupakan proporsi varians dalam variabel dependen yang diprediksi dari variabel independen [17]. Sementara, MAPE mengukur persentase kesalahan absolut rata-rata antara nilai prediksi dan nilai aktual, sehingga metrik ini efektif ketika variasi relatif lebih penting daripada variasi absolut [18]. Dengan menggunakan kedua metrik ini mampu untuk membandingkan model lebih komprehensif. Berdasarkan tabel 6, dapat dilihat bahwa model RFR menunjukkan performa yang terbaik. Hal ini karena nilai  $R^2$  yang paling tinggi dan MAPE paling kecil, walaupun nilai ini sebenarnya tidak berbeda jauh dengan model  $SARIMAX(2,0,1)(1,0,0)^{I2}$ . Artinya, kedua model ini mampu memprediksi produksi padi dengan hasil akurasi yang baik.



**Tabel 6.** Hasil Evaluasi Model

Model	$R^2$	MAPE (%)
SVR	0,914	12,090
RFR	0,976	7,827
SARIMAX(2,0,1)(1,0,0) <sup>12</sup>	0,962	16,398

Model RFR dapat memprediksi produksi padi melalui variabel luas panen, kelembaban (NDWI), suhu tanah, dan curah hujan. Pada konteks ini, RFR mampu menangani variabilitas dalam data iklim. Hal ini karena RFR membuat banyak pohon keputusan berdasarkan sampel acak dari *data training*, sehingga dapat menangkap pola dan variabilitas dalam data. Selain itu, model RFR ini dapat mengurangi pengaruh anomali dalam data karena hasil akhir menggunakan rata-rata dari prediksi semua pohon, sehingga apabila terdapat anomali pada satu pohon tidak akan berpengaruh pada keputusan akhir.

### III.2. Pembahasan

Dalam penelitian ini, model yang memberikan performa terbaik adalah model *Random Forest*. Model RFR memberikan  $R^2$  sebesar 0,976 yang berarti 97,6% variasi dari data dependen (produksi padi) dapat dijelaskan oleh variabel independen di dalam model. Hal ini membuktikan bahwa metode *machine learning* memiliki performa yang lebih akurat dalam memprediksi produksi padi di Jawa Tengah dibanding dengan metode *time series*. Sejalan dengan penelitian [20] yang menghasilkan prediksi yang lebih baik pada metode *machine learning* daripada metode *time series* (ARIMA). Prediksi yang akurat ini, tentu dapat mengoptimalkan ketahanan pangan di Jawa Tengah. Dalam konteks ini, prediksi dari model RFR ini dapat digunakan oleh pemerintah untuk membuat keputusan yang lebih tepat dalam mengelola sumber daya pertanian. Hal-hal yang bisa dilakukan pemerintah dalam bidang pertanian, sebagai berikut:

#### 1. Mengalokasikan Sumber Daya Pertanian secara Optimal

Berdasarkan temuan penelitian ini, luas panen memiliki pengaruh yang paling besar dalam memprediksi produksi padi, kemudian diikuti curah hujan. Sejalan dengan penelitian [21] bahwa luas panen dan curah hujan mempengaruhi produksi padi. Dari temuan ini, pemerintah dapat mengalokasikan lahan panen yang lebih luas untuk padi dan meningkatkan efisiensi penggunaan lahan agar menghasilkan produksi padi yang optimal. Selain itu, curah hujan juga berperan penting dalam produksi padi. Namun demikian, curah hujan tidak dapat dikontrol, hal yang bisa dilakukan adalah meningkatkan penggunaan teknologi irigasi dengan memanfaatkan air hujan. Salah satu teknologi yang dapat digunakan adalah teknologi panen hujan/*rainwater harvesting* [22]. Teknologi ini akan menampung air hujan sebagai air baku untuk sistem irigasi yang dapat dimanfaatkan pada musim kemarau, sehingga pada musim kemarau tetap menghasilkan produksi padi yang optimal.

#### 2. Program Bantuan Petani yang Tepat Sasaran

Hasil prediksi juga dapat memperkirakan stok padi yang dimiliki Jawa Tengah. Apabila dari hasil prediksi, stok padi di Jawa Tengah ternyata semakin berkurang. Hal ini dapat menjadi sinyal awal untuk pemerintah agar bersiap terlebih dahulu. Dengan ini, pemerintah dapat mengalokasikan bantuan secara tepat kepada petani yang produksi padinya masih kurang optimal. Bantuan dapat berupa edukasi teknologi irigasi, pupuk, benih, dan sebagainya. Dengan ini, diharapkan petani dapat menghasilkan produksi padi yang optimal sehingga kerawanan pangan dapat dicegah sedini mungkin.

#### 3. Memberikan Pengetahuan dan Keterampilan Petani



Walaupun hasil prediksi yang berpengaruh adalah luas panen dan curah hujan, pengetahuan dan keterampilan petani juga sangat penting untuk meningkatkan produksi padi. Hal ini dapat dilihat dari pola prediksi mendatang apabila juga terjadi penurunan produksi padi maka perhatian kepada para petani perlu ditingkatkan. Dalam konteks ini, pemerintah dapat meningkatkan pengetahuan dan keterampilan petani melalui program pelatihan dan penyuluhan. Program pelatihan dan penyuluhan dapat berupa mengenalkan praktik pertanian terbaru, teknologi pertanian yang lebih modern, dan manajemen sumber daya yang dapat membantu petani dalam meningkatkan produksi padi.

Dari poin-poin di atas, pemerintah dapat mengaplikasikan kepada para petani di Jawa Tengah agar produksi padi semakin meningkat. Dengan menerapkan strategi yang didasarkan pada hasil prediksi tersebut, diharapkan produksi padi di Jawa Tengah dapat meningkat secara signifikan. Apabila prediksi ini terbukti berhasil, maka Jawa Tengah menjadi pemasok padi yang paling besar di Indonesia. Dengan produksi padi yang meningkat ini, Jawa Tengah juga dapat menghasilkan beras yang optimal, sehingga dapat memenuhi kebutuhan pangan di Jawa Tengah, bahkan seluruh Indonesia. Hal ini juga dapat mengurangi ketergantungan Indonesia untuk mengimpor beras dari luar negeri. Pada akhirnya, hal ini akan berdampak pada peningkatan ketahanan pangan dan pertumbuhan ekonomi dalam negeri.

#### IV. KESIMPULAN

Berdasarkan dari hasil penelitian yang telah dilakukan terhadap ketiga model yang digunakan yaitu model SVR, RFR dan SARIMAX dengan menggunakan luas panen, *Normalized Difference Water Index (NDWI)*, suhu permukaan tanah, dan curah hujan dalam melakukan prediksi produksi padi di Jawa Tengah tahun 2018-2023 diperoleh RFR adalah model terbaik karena memiliki nilai R-squared yang tinggi dan nilai MAPE yang rendah dibandingkan dengan model yang lain, yakni 97,6% dan 7,827. Hal ini membuktikan keunggulan metode *machine learning* dibandingkan dengan metode time series dalam prediksi produksi padi. Prediksi yang akurat dari model ini berpotensi mengoptimalkan ketahanan pangan di Jawa Tengah dengan memberikan dasar yang kuat bagi pemerintah dan pemangku kepentingan untuk membuat keputusan yang lebih tepat dalam mengelola sumber daya pertanian. Pemerintah dapat mengalokasikan lahan yang lebih luas untuk padi dan meningkatkan efisiensi penggunaan lahan. Selain itu, dengan teknologi irigasi seperti *rainwater harvesting*, air hujan dapat dimanfaatkan secara efektif untuk mempertahankan produksi padi optimal selama musim kemarau. Prediksi produksi padi juga memungkinkan pemerintah untuk mengantisipasi potensi kerawanan pangan dan mengalokasikan bantuan secara tepat kepada petani yang membutuhkan, sehingga ketahanan pangan dapat terjaga dengan baik.

#### REFERENSI

1. A. M. Thow, S. K. Sharma, and C. N. Rachmi, “An analysis of Indonesia’s shrinking food security policy space under the WTO,” *Food Secur.*, vol. 11, no. 6, pp. 1275–1287, Dec. 2019, doi: 10.1007/s12571-019-00967-2.
2. J. D. B. Gil, P. Reidsma, K. Giller, L. Todman, A. Whitmore, and M. van Ittersum, “Sustainable development goal 2: Improved targets and indicators for agriculture and food security,” *Ambio*, vol. 48, no. 7, pp. 685–698, 2019, doi: 10.1007/s13280-018-1101-4.
3. Z. Chu and J. Yu, “An end-to-end model for rice yield prediction using deep learning fusion,” *Comput. Electron. Agric.*, vol. 174, no. April, p. 105471, 2020, doi: 10.1016/j.compag.2020.105471.
4. Y. Guo et al., “Integrated phenology and climate in rice yields prediction using machine learning methods,” *Ecol. Indic.*, vol. 120, p. 106935, 2021, doi: 10.1016/j.ecolind.2020.106935.



5. Y. Cai et al., “Agricultural and Forest Meteorology Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches,” *Agric. For. Meteorol.*, vol. 274, no. May 2018, pp. 144–159, 2019, doi: 10.1016/j.agrformet.2019.03.010.
6. S. Hama, J. Dash, and P. M. Atkinson, “Science of the Total Environment Forecasting wheat and barley crop production in arid and semi-arid regions using remotely sensed primary productivity and crop phenology : A case study in Iraq,” *Sci. Total Environ.*, vol. 613–614, pp. 250–262, 2018, doi: 10.1016/j.scitotenv.2017.09.057.
7. C. Folberth, A. Baklanov, J. Balkovi, R. Skalský, N. Khabarov, and M. Obersteiner, “Agricultural and Forest Meteorology Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning,” vol. 264, no. May 2018, pp. 1–15, 2019, doi: 10.1016/j.agrformet.2018.09.021.
8. S. H. Shah, Y. Angel, R. Houborg, S. Ali, and M. F. McCabe, “A Random Forest Machine Learning Approach for the Retrieval of Leaf Chlorophyll Content in Wheat,” 2019
9. J. Cao, Z. Zhang, F. Tao, L. Zhang, Y. Luo, and J. Zhang, “Agricultural and Forest Meteorology Integrating Multi-Source Data for Rice Yield Prediction across China using Machine Learning and Deep Learning Approaches,” vol. 297, no. April 2020, 2021.
10. BPS, “Luas Panen, Produksi, dan Produktivitas Padi Menurut Provinsi, 2018-2022.” [Online]. Available: [www.bps.go.id/id/statistics-table/2/MTQ5OCMy/luas-panen--produksi--dan-produktivitas-padi-menurut-provinsi.html](http://www.bps.go.id/id/statistics-table/2/MTQ5OCMy/luas-panen--produksi--dan-produktivitas-padi-menurut-provinsi.html)
11. P. H. Verburg and A. Veldkamp, “The role of spatially explicit models in land-use change research: A case study for cropping patterns in China,” *Agric. Ecosyst. Environ.*, vol. 85, no. 1–3, pp. 177–190, 2001, doi: 10.1016/S0167-8809(01)00184-0.
12. J. Wang, C. Wang, N. Chen, Z. Xiong, D. Wolfe, and J. Zou, “Response of rice production to elevated [CO<sub>2</sub>] and its interaction with rising temperature or nitrogen supply: a meta-analysis,” *Clim. Change*, vol. 130, no. 4, pp. 529–543, 2015, doi: 10.1007/s10584-015-1374-6.
13. W. J. Li et al., “Climate change impact and its contribution share to paddy rice production in Jiangxi, China,” *J. Integr. Agric.*, vol. 13, no. 7, pp. 1565–1574, 2014, doi: 10.1016/S2095-3119(14)60811-X.
14. N. Vaghefi, M. Nasir Shamsudin, A. Radam, and K. A. Rahim, “Modelling the impact of climate change on rice production: An overview,” *J. Appl. Sci.*, vol. 13, no. 24, pp. 5649–5660, 2013, doi: 10.3923/jas.2013.5649.5660.
15. K. Smets, B. Verdonk, and E. M. Jordaan, “Evaluation of performance measures for SVR hyperparameter selection,” *IEEE Int. Conf. Neural Networks - Conf. Proc.*, pp. 637–642, 2007, doi: 10.1109/IJCNN.2007.4371031.
16. M. Schonlau and R. Y. Zou, “The random forest algorithm for statistical learning,” *Stata J.*, vol. 20, no. 1, pp. 3–29, 2020, doi: 10.1177/1536867X20909688.
17. Wright, S., “Correlation and causation,” *Journal of agricultural research*, 20(7), 557, 1921.
18. De Myttenaere A, Golden B, Le Grand B, Rossi F., “Mean absolute percentage error for regression models,” *Neurocomputing* 192(1):38–48, 2016, doi 10.1016/j.neucom.2015.12.114.
19. M. Awad, “*Efficient Learning Machine: Theories, Concepts, and Applications for Engineers and System Designers*,” 2015. [Online]
20. Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. “A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks,” *Future Internet*, 15(8), 255, 2023, doi: 10.3390/fi15080255.
21. R. Alfred, J. O. E. H. Obit, and C. P. Chin, “Towards Paddy Rice Smart Farming : A Review on Big Data , Machine Learning , and Rice Production Tasks,” vol. 9, 2021, doi: 10.1109/ACCESS.2021.3069449.
22. T. T. Tolossa, F. B. Abebe, and A. A. Girma, “Review: Rainwater harvesting technology practices and implication of climate change characteristics in Eastern Ethiopia,” *Cogent Food Agric.*, vol. 6, no. 1, 2020, doi: 10.1080/23311932.2020.1724354.
23. C. Charatsari, E. D. Lioutas, and A. Koutsouris, “Farmer field schools and the co - creation of knowledge and innovation : the mediating role of social capital,” pp. 205–206, 2020.
24. L. Tang et al., “Food Security in China : A Brief View of Rice Production in,” 2022.
25. R. B. R. Firdaus et al., “Paddy , rice and food security in Malaysia : A review of climate change impacts GEOGRAPHY | REVIEW ARTICLE Paddy , rice and food security in Malaysia : A review of climate change impacts,” 2020, doi: 10.1080/23311886.2020.1818373.
26. Nodin, M. N., Mustafa, Z., & Hussain, S. I. “Assessing rice production efficiency for food security policy planning in Malaysia: A non-parametric bootstrap data envelopment analysis approach,”. *Food Policy*, 107, 102208, 2022, doi: 10.1016/j.foodpol.2021.102208.



**SENADA**  
Seminar Nasional Sains Data

Seminar Nasional Sains Data 2024 (SENADA 2024)  
UPN “Veteran” Jawa Timur

E-ISSN 2808-5841

P-ISSN 2808-7283

27. Connor, M., de Guia, A. H., Pustika, A. B., Sudarmaji, Kobarsih, M., & Hellin, J., “Rice farming in central Java, Indonesia—adoption of sustainable farming practices, impacts and implications,” *Agronomy*, *11*(5), 881, 2021, doi: 10.3390/agronomy11050881.