



Jurnal Teknologi Informasi dan Komunikasi

Vol: 14 No 02 2023

E-ISSN: 2477-3255

Diterima Redaksi: 24-09-2023 | Revisi: 20-10-2023 | Diterbitkan: 15-11-2023

Pemodelan Spasial untuk Analisa Produksi Padi Integrasi Machine Learning

Desy Ika Puspitasari¹, Al Fath Riza Kholdani², Tri Wahyu Qur'ana³, Adani Dharmawati⁴

^{1,2,3,4}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Islam Kalimantan
(UNISKA) MAB Banjarmasin

^{1,2,3,4}Jl. Adhyaksa No.2 Kayu Tangi Banjarmasin. Kode Post 70123. No Telp. (0511)3304029

e-mail: ¹smile4desyka@gmail.com, ²kholdanialfath@gmail.com, ³twqurana@gmail.com,

⁴adani.dharmawati@gmail.com

Abstrak

Pertumbuhan populasi manusia yang terus meningkat memicu minat dalam memprediksi produksi padi yang akurat. Penelitian ini memadukan Sistem Informasi Geografi dan machine learning untuk memodelkan prediksi produksi padi dengan kesalahan minimal. Menggunakan regresi linear, penelitian ini menganalisis produksi padi di 13 kabupaten/kota di Kalimantan Selatan dengan data spasial dan historis sebagai fitur, serta produksi padi sebagai variabel target. Hasil eksperimen menunjukkan bahwa pendekatan ini menghasilkan prediksi yang akurat, dengan Algoritma Linear Regression memiliki tingkat akurasi yang konsisten dan R^2 sebesar 0.992, menjelaskan bahwa 99.2% variasi variabel respon dapat dijelaskan oleh variabel predictor, dan 0.8% dipengaruhi faktor lain. Sementara Kabupaten Hulu Sungai Utara memiliki produksi padi tertinggi dan kabupaten Barito Kuala terendah. Studi ini memberikan kontribusi penting dalam pemodelan spasial untuk prediksi produksi padi, membantu para ahli pertanian dan pengambil keputusan mengidentifikasi potensi produksi padi serta meningkatkan akurasi prediksi produksi padi.

Kata kunci: Pemodelan spasial, Prediksi produksi padi, Sistem Informasi Geografi, Machine learning, Linear regression

Spatial Modeling for Rice Production Analysis Integrating Machine Learning

Abstract

The growing human population has sparked interest in accurate rice production prediction. This research combines Geographic Information Systems and machine learning to model rice paddy production prediction with minimal error. Using linear regression, this research analyzes rice paddy production in 13 districts/cities in South Kalimantan with spatial and historical data as features, and rice paddy production as the target variable. Experimental results show that this approach produces accurate predictions, with the Linear Regression Algorithm having a consistent level of accuracy and R^2 of 0.992, explaining that 99.2% of the variation in the response variable can be explained by the predictor variable, and 0.8% is influenced by other factors. While Hulu Sungai Utara district has the highest rice production and Barito Kuala district has the lowest. This study makes an important contribution in spatial modeling for rice paddy production prediction, helping agricultural experts and decision makers identify rice paddy production potential and improve the accuracy of rice paddy production prediction.

<https://doi.org/10.31849/digitalzone.v14i2.16256>

Digital Zone is licensed under a Creative Commons Attribution International (CC BY-SA 4.0)

Keywords: *Spatial modeling, Rice production prediction, Geographic Information System, Machine learning, Linear regression.*

1. Pendahuluan

Produksi padi di Kalimantan Selatan tahun 2021 tercatat sebesar 1.041.862,91 ton. Angka ini turun sebesar 9,42 persen dibandingkan produksi padi tahun 2020 sebesar 1.150.306,66 ton. Penurunan produksi tersebut seiring dengan menurunnya luas panen sebesar 11,75 persen [1]. Dalam hal ini, produksi padi sebagai sumber makanan pokok bagi sebagian besar populasi dunia memegang peranan yang sangat penting. Namun, dalam menghadapi pertumbuhan populasi yang cepat dan perubahan lingkungan yang terjadi, tantangan untuk memprediksi produksi padi secara akurat menjadi semakin kompleks.

Pada saat ini, pemodelan spasial dan penggunaan teknik *machine learning* telah membuka peluang baru dalam menganalisis dan memprediksi produksi padi dengan tingkat akurasi yang lebih tinggi [2]–[6]. Pemodelan spasial memungkinkan integrasi data spasial untuk memahami hubungan spasial yang kompleks dalam produksi padi. Sementara itu, *machine learning* memberikan kemampuan untuk melatih model dengan menggunakan data historis dan mengenali pola yang tersembunyi dalam data untuk memprediksi produksi padi [7].

Namun, walaupun ada kemajuan signifikan dalam pemodelan spasial dan penggunaan *machine learning*, masih terdapat tantangan yang perlu diatasi. Salah satu tantangan utama adalah mengintegrasikan pemodelan spasial dengan teknik *machine learning* secara efektif. Selain itu, penting juga untuk memilih algoritma *machine learning* yang tepat, seperti regresi linier, yang dapat mengungkapkan hubungan linier antara variabel spasial dan produksi padi. Dalam rangka mengatasi tantangan ini, penelitian ini bertujuan untuk melakukan pemodelan spasial untuk analisis produksi padi dengan integrasi *machine learning* menggunakan regresi linier, dengan fokus pada penggunaan sistem informasi geografi (SIG) sebagai kerangka kerjanya. Diharapkan dengan pendekatan ini, prediksi produksi padi dapat dilakukan dengan tingkat akurasi yang lebih tinggi, yang akan memberikan manfaat besar bagi para ahli pertanian dan pengambil keputusan dalam perencanaan dan pengelolaan produksi padi.

Regresi linier bekerja dengan baik ketika hubungan antara variabel independen dan dependen dapat dijelaskan dengan hubungan linier [13]. Dalam kasus tersebut, regresi linier dapat memberikan hasil yang cukup akurat. Regresi linier juga stabil dalam situasi dengan jumlah sampel kecil.

Model prediksi produksi padi menggunakan *machine learning* sudah di review pada penelitian sebelumnya [8]–[13] sebagian besar menekankan pada prediksi produksi padi dengan algoritma *machine learning*. Penelitian terkait prediksi terhadap produksi padi menggunakan algoritma backpropagation neural network [10], [12], metode klasifikasi data mining untuk prediksi [9], algoritma Random forest [13], dan K-Nearest Neighbor [8]. Ada juga penelitian yang sama menerapkan algoritma Linear Regression untuk prediksi padi [11] namun tidak disertai analisa spasial dan penjelasan hubungan antar variabel.

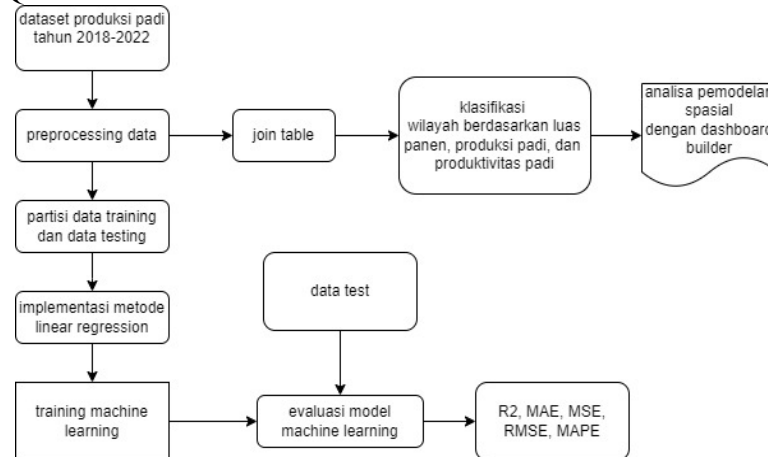
Berdasarkan latar belakang masalah, tujuan dari penelitian ini ialah memperoleh pemodelan spasial untuk analisa produksi padi integrasi *machine learning*. Penelitian sebelumnya kurang menganalisis hubungan antar variabel dan pemodelan spasial. Padahal, analisis hubungan antar variabel dapat membantu meningkatkan akurasi prediksi. Penelitian ini dapat mengetahui prediksi produktivitas padi di Kalimantan Selatan dengan menggunakan dengan menggunakan *machine learning*, linear regression (LR) untuk mendapatkan nilai error terkecil antara data aktual dan data prediksi, sehingga dapat membantu memberi masukan untuk mengembangkan kebijakan pertanian dan pangan yang efektif pada skala regional dan global, mengingat sangat pentingnya keakuratan prediksi produktivitas padi.

2. Metode Penelitian

Dalam riset ini, dataset produksi padi dari tahun 2018 hingga tahun 2022 telah melalui tahap *preprocessing* data untuk melakukan penggabungan data spasial dan data atribut. Setelah

preprocessing data selesai, penggunaan *machine learning* di platform Knime membagi data menjadi data training (70%) dan data testing (30%). Hasil evaluasi kinerja model, baik pada data training maupun data testing, mencakup nilai R^2 yang mendekati 1, MAE (*Mean Absolute Error*), MSE (*Mean Squared Error*), RMSE (*Root Mean Squared Error*), dan MAPE (*Mean Absolute Percentage Error*) dengan kesalahan terendah.

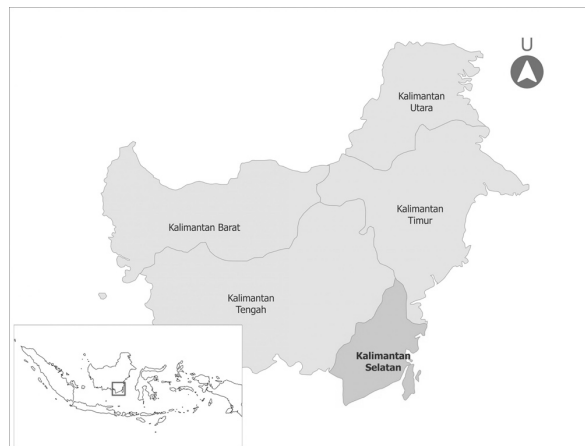
Selanjutnya, pada tahap pemetaan menggunakan QGIS, wilayah-wilayah diklasifikasikan berdasarkan range produksi padi. Analisis pemodelan spasial dilakukan melalui *dashboard builder* plugin QGIS.



Gambar 1. Pemodelan Spasial Prediksi Produksi Padi Integrasi Machine Learning

2.1. Area Studi

Wilayah studi dari penelitian ini adalah Provinsi Kalimantan Selatan, Indonesia seperti yang ditunjukkan pada gambar diatas. Daerah penelitian dipilih provinsi Kalimantan Selatan karena sebagai sentra produksi padi di Kalimantan [14].



Gambar 2. Area Studi Penelitian (Provinsi Kalimantan Selatan)

2.2. Sumber Data

1. Data Spasial : peta Provinsi Kalimantan Selatan beserta batas administrasinya diambil dari sumber : <https://geosai.my.id/download-shp-kabupaten-kota-indonesia/>
2. Data atribut : berupa Luas Panen, Produksi dan Produktivitas, sumber dari website resmi Badan Pusat Statistika (BPS), Survei Kerangka Sampel Area (KSA)/BPS-Statistics Indonesia, Area Sampling Frame (ASF) Survey.

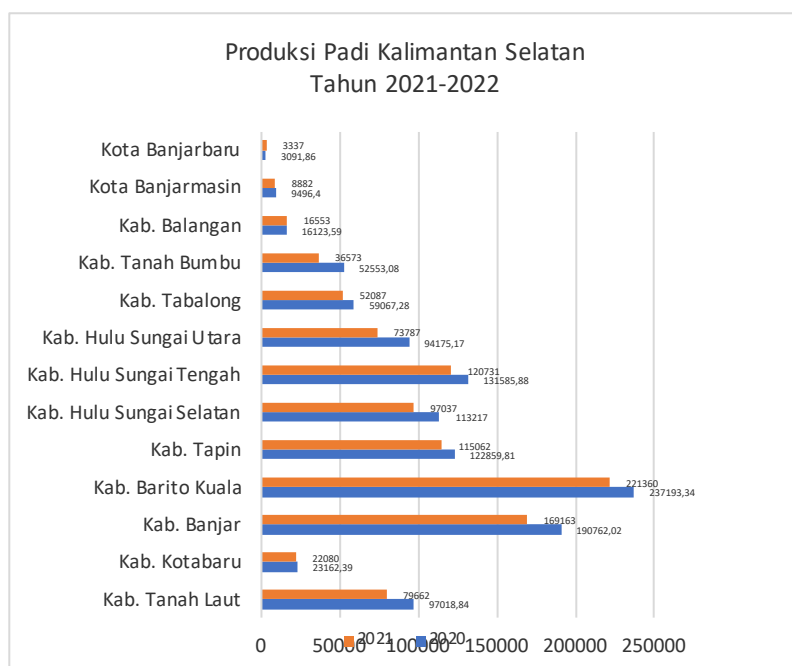
Adapun kabupaten yang dipilih berdasarkan banyaknya jumlah sampel KSA padi di Kalimantan Selatan sehingga menghasilkan kabupaten amatan pada Tabel 1.

Tabel 1. Kabupaten Penelitian

Kabupaten/Kota	2020	2021
Kab. Tanah Laut	97018.84	79662
Kab. Kotabaru	23162.39	22080
Kab. Banjar	190762.02	169163
Kab. Barito Kuala	237193.34	221360
Kab. Tapin	122859.81	115062
Kab. Hulu Sungai Selatan	113217	97037
Kab. Hulu Sungai Tengah	131585.88	120731
Kab. Hulu Sungai Utara	94175.17	73787
Kab. Tabalong	59067.28	52087
Kab. Tanah Bumbu	52553.08	36573
Kab. Balangan	16123.59	16553
Kota Banjarmasin	9496.4	8882
Kota Banjarbaru	3091.86	3337

Catatan : Produksi Padi (Ton)

*Sumber : Dinas Tanaman Pangan dan Hortikultura Provinsi Kalimantan Selatan per 12 September 2022



Gambar 3. Grafik Produksi padi th 2020-2021 Pro. Kalimantan Selatan

2.3. Preprocessing Data pada QGIS

Proses pengolahan data pada QGIS dilakukan sebelum tahap penggabungan data spasial dan data atribut, dimana penting untuk memastikan bahwa tipe data pada data atribut sesuai agar *join table* dapat dilakukan. Pada tahap ini, langkah pertama adalah menjalankan proses refactor fields pada Gambar 4.

Gambar 4 menunjukkan adanya ketidaksesuaian tipe data, khususnya terdapat tipe data teks pada atribut "Luas Panen_2018" dan "Produksi_2018," sedangkan tipe data yang diperlukan di

Gambar 6 diatas merupakan desain eksperimen menggunakan layout node repository dengan metode linear regression. Pengujian dilakukan dengan menggunakan matriks *linear correlation* dan hasilnya adalah output nilai akurasi prediksi yang terdiri dari skor R^2 , MSE, MAE dan RMSE dalam bentuk *numeric score*. *Math* formula digunakan untuk mengetahui presentase gap data aktual dan data prediksi. Selain itu, scatter plot juga digunakan untuk mengetahui korelasi antara variabel prediksi produksi padi.

3. Hasil dan Pembahasan

Adapun kabupaten yang dipilih berdasarkan banyaknya jumlah sampel KSA padi di 13 kabupaten/kota provinsi Kalimantan Selatan tahun 2018-2022, sehingga data yang digunakan tampak pada tabel 1 berikut :

Tabel 2. Variabel Penelitian

Data	Variabel
Produksi padi	Dependen/respon (y)
Luas panen	Independen/predictor (x_1)
Produktivitas	Independen/predictor (x_2)

Secara umum, algoritma linear regression dibagi menjadi dua jenis, yaitu *simple linear regression* dan *multiple linear regression*. *Simple linear regression* merupakan hubungan antara satu variabel dependen dengan satu variabel independen, sedangkan *multiple linear regression* merupakan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen.

Dalam regresi linear, variabel respon merupakan variabel dependen, sedangkan variabel independen adalah variabel yang digunakan untuk memprediksi (*variable predictor*), yakni menjelaskan variasi dalam variabel respon. Dalam kajian prediksi produksi padi, variabel respon berupa jumlah produksi padi dalam satuan tertentu (ton/GKG). Perhitungan yang digunakan untuk *multiple linear regression* dinyatakan dalam (1).

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

dengan

Y = variabel dependen atau variabel tak bebas (nilai yang diprediksikan)

$X_1, X_2, \dots, X_n$ = variabel independen atau variabel bebas

a = konstanta

$b_1, b_2, \dots, b_n$ = koefisien regresi.

Dengan mengacu pada (1), untuk menghitung produksi padi dengan algoritma linear regression, digunakan (2).

$$Y = a + b_1X_1 + b_2X_2 \quad (2)$$

dengan

Y = produksi padi

a = konstanta

b_1 = koefisien variabel Luas panen

X_1 = Luas Panen

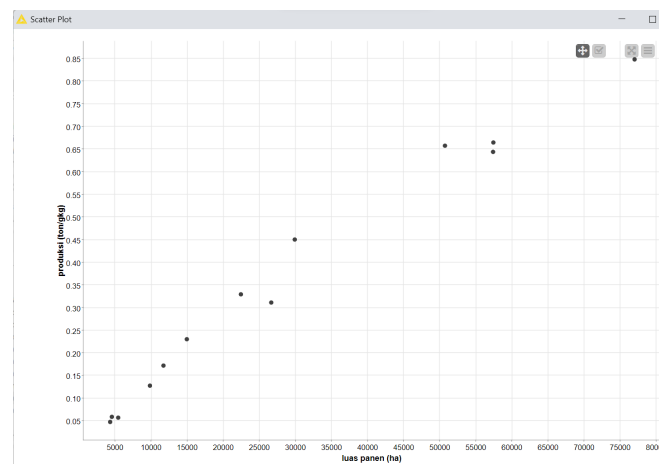
b_2 = koefisien variabel Produktivitas

X_2 = Produktivitas

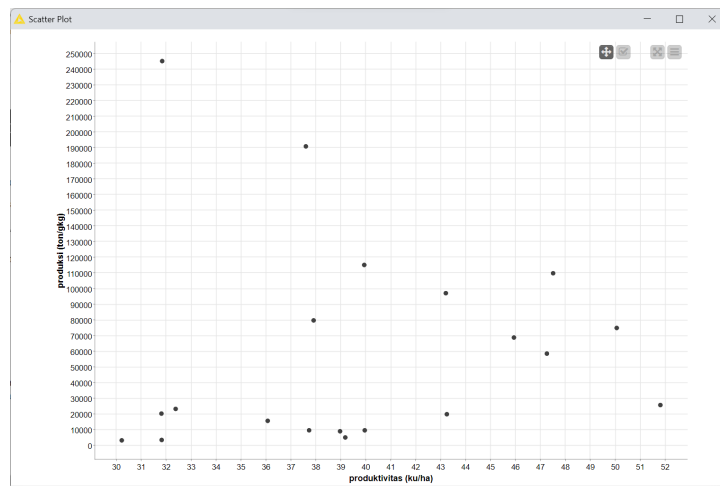
3.1. Pola hubungan Variabel Respon dan Prediktor

Berdasarkan scatter plot yang terlihat pada Gambar 7, dapat disimpulkan bahwa hubungan antara produksi padi dan luas panen masuk dalam kategori komponen parametrik. Ini berarti bahwa hubungan ini didasarkan pada pendekatan dengan variabel data yang memiliki pola yang diketahui atau dapat diprediksi.

Sementara pada Gambar 8 diatas adalah bagian dari komponen nonparametrik, yang berarti bahwa pendekatan ini menggunakan variabel data yang polanya tidak diketahui dan dapat membentuk kurva sendiri saat dianalisis [15]

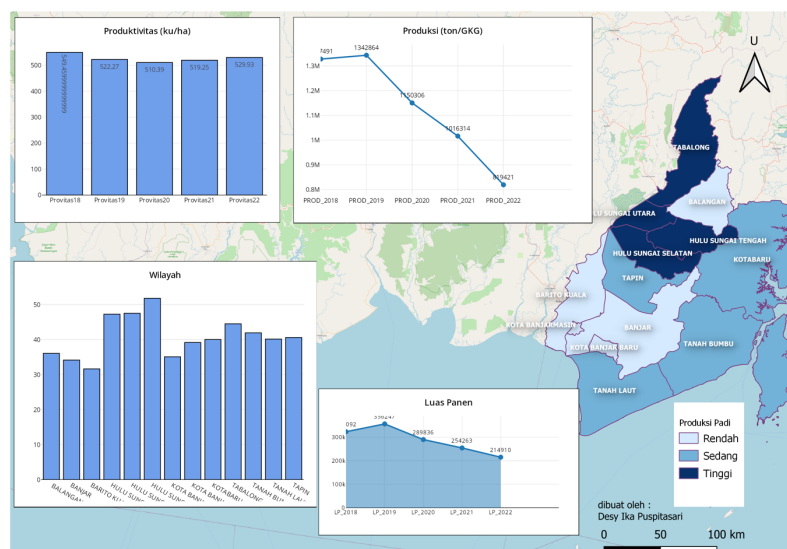


Gambar 7. Scatter plot Luas Panen vs Produksi



Gambar 8. Scatter plot Produktivitas vs Produksi

3.1. Hasil Pemodelan Spasial

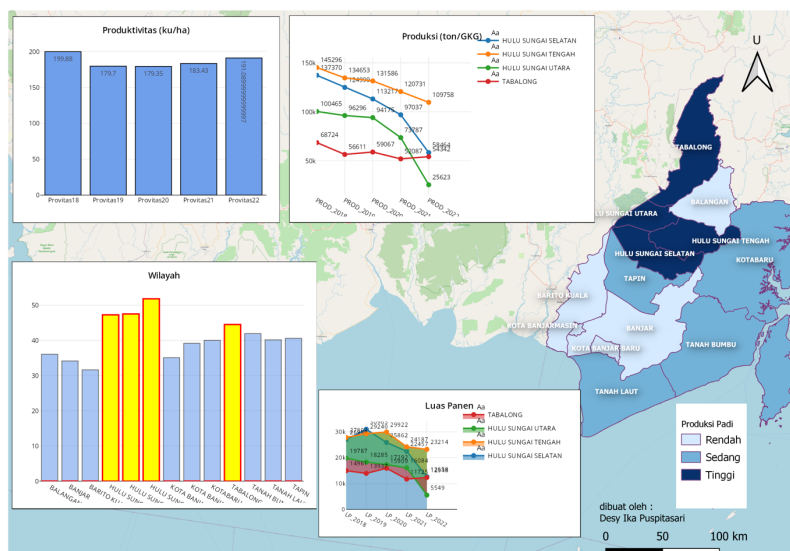


Gambar 9. Produksi padi di 13 wilayah Provinsi Kalimantan Selatan dari tahun 2018-2022

<https://doi.org/10.31849/digitalzone.v14i2.16256>

Digital Zone is licensed under a Creative Commons Attribution International (CC BY-SA 4.0)

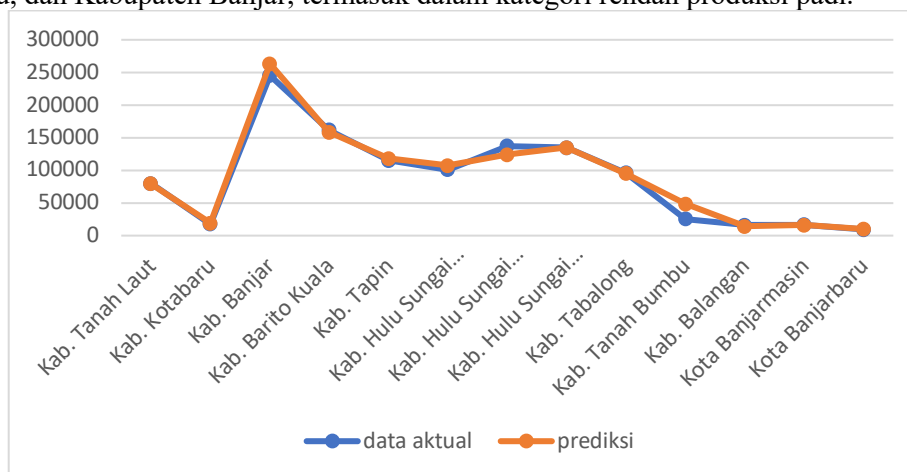
Gambar 9 adalah visualisasi pemodelan spasial di 13 wilayah kabupaten/kota Provinsi Kalimantan Selatan. Dari grafik Produksi, terlihat penurunan produksi padi dari tahun 2018 hingga 2022. Wilayah dengan produksi padi tertinggi adalah Kabupaten Hulu Sungai Utara, sementara Kabupaten Barito Kuala memiliki produksi padi terendah.



Gambar 10. Pemodelan spasial wilayah *prediksi* produksi padi paling tinggi di Provinsi Kalimantan Selatan

Pada Gambar 10, terdapat visualisasi produksi padi dengan klasifikasi berdasarkan tiga kategori nilai, yaitu rendah, sedang, dan tinggi. Dalam kategori tersebut, terdapat 5 wilayah dengan produksi padi rendah, 4 wilayah dengan produksi padi sedang, dan 4 wilayah dengan produksi padi tinggi.

Wilayah yang masuk dalam kategori tinggi produksi padi adalah Hulu Sungai Utara, Hulu Sungai Tengah, Hulu Sungai Selatan, dan Tabalong. Sedangkan wilayah yang masuk dalam kategori sedang produksi padi adalah Tapin, Tanah Bumbu, Kotabaru, dan Tanah Laut. Wilayah-wilayah lainnya, yaitu Kabupaten Balangan, Barito Kuala, Kota Banjarmasin, Kota Banjarbaru, dan Kabupaten Banjar, termasuk dalam kategori rendah produksi padi.



Gambar 11. Perbandingan data Aktual dengan hasil Prediksi Produksi Padi

Berdasarkan grafik perbandingan antara data aktual dan hasil prediksi menggunakan math formula pada Knime, yang ditampilkan pada Gambar 11, dapat disimpulkan bahwa variabel respon aktual dan data prediksi memiliki pola yang serupa. Ini mengindikasikan bahwa data prediksi secara efektif mengikuti data aktual, menunjukkan bahwa pemodelan yang digunakan telah sesuai dan mampu memprediksi produksi padi dengan baik.

Tabel 3. Hasil Akurasi Prediksi Produksi Padi menggunakan Linear Regression

Keterangan	Nilai
R^2	0.992
Mean Absolute Error	0.02
Mean Squared Error	0.001
Root Mean Squared Error	0.027
Mean Signed Difference	-0.011
Mean Absolute Percentage Error	0.367

Terlihat dari hasil akurasi yang tercantum dalam Tabel 3, algoritma Linear Regression yang diuji menunjukkan kinerja yang memuaskan dengan tingkat akurasi yang konsisten. Dengan nilai koefisien determinasi (R^2) sebesar 0.992, MSE sebesar 0.001, RMSE sebesar 0.027, dan MAPE sebesar 0.367, dapat disimpulkan bahwa semakin tinggi nilai koefisien determinasi, semakin baik performa model. Selain itu, semakin kecil nilai MSE, RMSE, dan MAPE, semakin baik pula model tersebut. Nilai akurasi keseluruhan penelitian ini lebih baik daripada penelitian Masdian, dkk. [13] yang memiliki akurasi 91%. Namun, keduanya menggunakan metode yang sama yaitu linear regression, serta memiliki akurasi yang tinggi.

4. Kesimpulan

Hasil analisis pemodelan spasial menunjukkan bahwa wilayah dengan produksi padi tertinggi terletak di Kabupaten Hulu Sungai Utara, sementara wilayah dengan produksi padi terendah berada di Kabupaten Barito Kuala. Rancangan prediksi produksi padi dengan model regresi linear menghasilkan nilai koefisien determinasi R^2 sebesar 0,992. Hal ini mengindikasikan bahwa sekitar 99,2% dari variabel prediktor dapat menjelaskan variasi dalam variabel respon, sedangkan sisanya sekitar 0,8% dipengaruhi oleh faktor lain.

Namun, perlu dicatat bahwa regresi linear memiliki beberapa asumsi yang harus dipenuhi, termasuk asumsi tentang distribusi dan hubungan linier antara variabel. Jika asumsi-asumsi ini tidak terpenuhi, maka metode regresi linear mungkin tidak akan menghasilkan hasil yang akurat. Dalam situasi seperti itu, mungkin perlu mempertimbangkan penggunaan metode lain yang lebih kompleks atau fleksibel. Oleh karena itu, penelitian berikutnya akan mempertimbangkan faktor-faktor lain yang dapat mempengaruhi prediksi produksi padi.

Daftar Pustaka

- [1] Badan Pusat Statistik (BPS), "Provinsi Kalimantan Selatan Dalam Angka 2022," Kalimantan Selatan, 2022. Accessed: Nov. 27, 2022. [Online]. Available: kalsel.bps.go.id
- [2] K. Kopczewska, "Spatial machine learning: new opportunities for regional science," *Annals of Regional Science*, vol. 68, no. 3, pp. 713–755, Jun. 2022, <http://dx.doi.org/10.1007/s00168-021-01101-x>.
- [3] A. Karim, D. S. Sarra, R. Wasono, T. W. Utami, and Toheri, "Spatial modelling for rice production analysis in Central Java province Indonesia," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2019. <http://dx.doi.org/10.1088/1742-6596/1217/1/012113>.
- [4] Y. Casali, N. Y. Aydin, and T. Comes, "Machine learning for spatial analyses in urban areas: a scoping review," *Sustainable Cities and Society*, vol. 85. Elsevier Ltd, Oct. 01, 2022. <http://dx.doi.org/10.1016/j.scs.2022.104050>.
- [5] A. Toomanian, A. A. Kakroodi, and M. A. Etemadi, "Spatial Modeling of Land Subsidence Using GIS-based Machine Learning Algorithms," *SSRN Electronic Journal Elsevier BV*, 2022.
- [6] B. Liang *et al.*, "Analysing and simulating spatial patterns of crop yield in Guizhou Province based on artificial neural networks," *Prog Phys Geogr*, vol. 45, no. 1, pp. 33–52, Feb. 2021, <http://dx.doi.org/10.1177/0309133320956631>.

<https://doi.org/10.31849/digitalzone.v14i2.16256>

-
- [7] A. Wilson, R. Sukumar, and N. Hemalatha, "Machine Learning Model for Rice Yield Prediction Using KNN Regression." *agriRxiv*, 2021. <http://dx.doi.org/10.31220/agriRxiv.2021.00070>.
- [8] W. Tb Panjaitan, E. Utami, and H. Al Fatta, "Prediksi Panen Padi Menggunakan Algoritma K-Nearest Neighbour," in *Prosiding SNATIF ke-5 Tahun 2018*, Fakultas Teknik - Universitas Muria Kudus, 2018, pp. 621–628. [Online]. Available: <https://yogyakarta.bps.go.id>.
- [9] S. Maesaroh, "Sistem Prediksi Produktifitas Pertanian Padi Menggunakan Data Mining," vol. 7, no. 2, 2017.
- [10] G. Ramadhona, B. Darma Setiawan, and F. A. Bachtiar, "Prediksi Produktivitas Padi Menggunakan Jaringan Syaraf Tiruan Backpropagation," 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, "Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi," *JNTETI*, vol. 8, no. 4, 2019.
- [12] H. Putra and N. Ulfa Walmi, "Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 2, pp. 100–107, Sep. 2020, <http://dx.doi.org/10.25077/teknosi.v6i2.2020.100-107>.
- [13] A. Raka Masdian, N. Bashit, and F. Hadi, "Analisis Produktivitas Padi Menggunakan Algoritma Machine Learning Random Forest di Kabupaten Batang Tahun 2018-2022," 2023.
- [14] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, Dec. 2020, <http://dx.doi.org/10.38094/jastt1457>.
- [15] M. Alkaff, H. Khatimi, W. Puspita, and Y. Sari, "Modelling and predicting wetland rice production using support vector regression," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 2, pp. 819–825, 2019, <http://dx.doi.org/10.12928/TELKOMNIKA.V17I2.10145>.
- [16] T. W. Utami, I. Manfaati Nur, and E. Suryaningsih, "Pemodelan Produksi Padi Menggunakan Regresi Semiparametrik Kernel," *Jurnal Statistika*, vol. 6, no. 2, pp. 160–165, Nov. 2018, [Online]. <https://doi.org/10.26714/jsunimus.6.2.2018.%25p>
-