



SISTEM PREDIKSI PRODUKSI PADI DI SUMATERA MENGGUNAKAN REGRESI LINEAR

Ery Permana Yudha^{1*}, Arif Rohmadi², Agung Teguh Setyadi³

^{1,2}Program Studi Informatika, Universitas Sebelas Maret, ³Program Studi D3 Teknik Informatika, Politeknik Elektronika Negeri Surabaya

Jln. Ir. Sutami No.36, Kentingan, Jebres, Surakarta, Jawa Tengah 57126

^{1*}erypermana@staff.uns.ac.id, ²arifrohmedi@staff.uns.ac.id, ³agungteguhsetyadi@pens.ac.id

Abstract

Sumatra is one of the islands that is a national rice barn because it is one of the largest rice producing areas in Indonesia. However, high productivity on the island of Sumatra also has several challenges such as uncertain climate change, land area, rainfall, humidity, and average temperature. To overcome these problems, innovative and data-based strategies are needed. One of these strategies is by implementing data processing to produce a rice productivity prediction model. This technique involves algorithms and machine learning to analyze patterns and trends in agriculture. Where the model is able to facilitate stakeholders concerned to prepare national food needs so that they are always met. In this study, a method of predicting rice productivity in Sumatra is proposed using the linear regression method. This study produces prediction models for each province in Sumatra. In general, the stages carried out are preprocessing, feature selection, training and testing, and evaluation. The trial was carried out by calculating the Mean Squared Error (MSE) value. Several algorithms, namely Linear Regression, Support Vector Regression (SVR), Random Forest Regression (RFR) produced an average MSE value of 0.022; 0.075; 0.026. Linear regression is able to produce a better model compared to the SVR and RFR methods.

Keywords : Prediction, Linear Regression, Support Vector Machine, Random Forest Regression, Mean Squared Error

Abstrak

Pulau Sumatera merupakan salah satu pulau yang menjadi lumbung padi nasional karena sebagai salah satu daerah penghasil padi terbesar di Indonesia. Namun, produktivitas yang tinggi di pulau Sumatera juga terdapat beberapa tantangan seperti perubahan iklim yang tidak menentu, luas lahan, curah hujan, kelembapan, dan suhu rata-rata. Untuk mengatasi permasalahan tersebut perlu strategi yang inovatif dan berbasis data. Salah satu strategi tersebut dengan menerapkan pengolahan data untuk menghasilkan model prediksi produktivitas padi. Teknik ini melibatkan algoritma dan pembelajaran mesin untuk menganalisis pola dan tren dalam pertanian. Model ini mempermudah stakeholder terkait untuk mempersiapkan kebutuhan pangan nasional agar selalu terpenuhi. Pada penelitian ini, diusulkan sebuah metode prediksi produktivitas padi di Sumatera menggunakan metode regresi linear. Penelitian ini menghasilkan model prediksi masing-masing di setiap provinsi di Sumatera. Secara umum, tahapan yang dilakukan yaitu preprocessing, seleksi fitur, training dan testing, dan evaluasi. Uji coba yang dilakukan dengan menghitung nilai Mean Squared Error (MSE). Beberapa algoritma yaitu Regresi Linear, Support Vector Regression (SVR), Random Forest Regression (RFR) menghasilkan nilai rata-rata MSE sebesar 0,022; 0,075; 0,026. Regresi linear mampu menghasilkan model yang lebih baik dibandingkan metode SVR dan RFR.

Kata kunci : Prediksi, Regresi Linear, Support Vector Machine, Random Forest Regression, Mean Squared Error



1. PENDAHULUAN

Produksi padi di Indonesia merupakan tulang punggung ketahanan pangan nasional. Pulau Sumatera, sebagai salah satu lumbung padi nasional, memiliki peran penting dalam memenuhi kebutuhan pangan domestik [1]. Sebagai salah satu daerah penghasil padi terbesar di Indonesia, Sumatera menghadapi berbagai tantangan yang dapat mempengaruhi produktivitas pertaniannya [2]. Tantangan ini mencakup perubahan iklim yang tidak menentu, luas lahan, curah hujan, kelembapan, suhu rata-rata. Untuk mengatasi permasalahan ini, diperlukan strategi yang inovatif dan berbasis data untuk meningkatkan efisiensi dan produktivitas pertanian.

Analisis prediksi menawarkan solusi yang menjanjikan dalam konteks ini. Dengan menggunakan teknologi ini, petani dan pengambil kebijakan dapat memanfaatkan data historis dan real-time untuk memprediksi hasil produksi padi [3]. Teknik ini melibatkan penggunaan algoritma statistik dan pembelajaran mesin untuk menganalisis pola dan tren dalam data pertanian, seperti luas lahan, curah hujan, kelembapan, suhu rata-rata. Hasil analisis ini dapat memberikan wawasan yang lebih akurat mengenai waktu tanam yang optimal, kebutuhan irigasi, dan strategi pengendalian hama [4], [5].

Implementasi analisis prediksi di Sumatera dapat memberikan berbagai manfaat, termasuk peningkatan produktivitas, pengurangan risiko gagal panen, dan efisiensi penggunaan sumber daya [6]. Misalnya, dengan memprediksi pola cuaca, petani dapat menyesuaikan waktu tanam dan panen untuk memaksimalkan hasil. Selain itu, analisis prediksi juga dapat membantu dalam perencanaan distribusi hasil panen, sehingga dapat mengurangi kehilangan pasca panen dan memastikan pasokan beras yang stabil.

Namun, penerapan teknologi ini tidak tanpa tantangan. Keterbatasan infrastruktur teknologi di beberapa daerah, kurangnya akses terhadap data berkualitas, serta kebutuhan akan pelatihan bagi petani dalam menggunakan teknologi baru merupakan beberapa hambatan yang perlu diatasi [7]. Oleh karena itu, kolaborasi antara ahli di bidang pertanian dan bidang teknologi informasi khususnya data sains diperlukan untuk menciptakan sebuah model yang mampu memprediksi produktivitas tanaman [8], [9].

Prediksi produktivitas pertanian hadir sebagai metode yang penting untuk meningkatkan efisiensi dan keberlanjutan produksi pangan. Dengan kemajuan teknologi dan analisis data, prediksi produktivitas pertanian dapat memberikan berbagai manfaat yang signifikan. Salah satu manfaat utama dari prediksi produktivitas adalah kemampuan untuk merencanakan dan mengelola sumber daya secara lebih efisien. Petani dapat memanfaatkan prediksi untuk menentukan waktu tanam yang optimal, memilih varietas tanaman yang sesuai, dan mengatur penggunaan input seperti pupuk dan pestisida. Hal ini tidak hanya meningkatkan hasil panen tetapi juga mengurangi dampak negatif terhadap lingkungan.

Beberapa contoh penelitian yang sudah pernah dilakukan seperti menggunakan teknologi penginderaan jarak jauh dan pembelajaran mesin untuk memprediksi hasil panen di berbagai wilayah [10], [11] mengintegrasikan data satelit dengan informasi cuaca dan kondisi tanah untuk memprediksi produktivitas padi [12], dan prediksi hasil pertanian berdasarkan perkembangan lahan pertanian [13]. Akan tetapi pada penelitian sebelumnya hanya dilakukan terhadap satu provinsi atau satu kabupaten saja.

Maka dari itu, pada penelitian ini kami mengusulkan sebuah sistem prediksi produksi padi di seluruh provinsi yang ada di Sumatera menggunakan metode regresi linear. Penelitian ini berkontribusi penting dalam bidang teknologi informasi dan pertanian. Hasil yang didapatkan pada penelitian ini berupa prediksi jumlah produksi padi berdasarkan beberapa parameter yang dipertimbangkan.

2. TINJAUAN PUSTAKA

Berdasarkan studi literatur yang terkait dengan penelitian serupa seperti analisis produksi padi di Indonesia [14], [15]. Penelitian tersebut menganalisis faktor-faktor yang mempengaruhi produksi padi seperti luas panen, tenaga kerja, dan investasi terhadap produksi padi di Indonesia. Metode penelitian dilakukan dengan pendekatan kuantitatif menggunakan data sekunder dan primer.

Beberapa penelitian terkini dilakukan uji coba dengan metode multicollinearity dan autokorelasi dari seluruh faktor. Hasil dari penelitian itu menunjukkan seluruh faktor tersebut sangat berpengaruh terhadap produksi



padi dan mencapai tingkat determinasi sebesar 0,983 [14]. Selain itu, metode multicollinearity sangat handal untuk menganalisis hubungan setiap faktor.

Penelitian terkini untuk memprediksi produktivitas padi menggunakan metode regresi linear pernah dilakukan di beberapa kabupaten [16][17][18]. Pada penelitian tersebut menunjukkan metode regresi linear sangat handal dalam memprediksi produktivitas padi. Metode regresi linear dapat memahami dan menentukan berbagai faktor yang berpengaruh terhadap produksi padi, faktor tersebut seperti luas tanam, luas panen, curah hujan, dan puso. Bahkan, memungkinkan *stakeholder* untuk mengambil keputusan yang lebih efisien berdasarkan informasi tersebut.

Penelitian serupa juga telah dilakukan seperti analisis regresi linear untuk memprediksi produktivitas sebuah tanaman kelapa sawit [19]. Pada penelitian tersebut dilakukan untuk menganalisis dan mengestimasi produktivitas kelapa sawit. Metode regresi linear digunakan untuk mengidentifikasi dan mengukur pengaruh variabel-variabel terhadap produktivitas kelapa sawit. Variabel-variabel yang digunakan seperti luas tanaman dalam hektar, curah hujan, dan jumlah produksi dalam ton. Berdasarkan hasil uji coba, penelitian tersebut mampu memprediksi produktivitas kelapa sawit dengan sangat baik. Selain itu, mampu mengidentifikasi faktor yang relevan terhadap produktivitas kelapa sawit.

Kemudian penelitian selanjutnya menganalisis korelasi antar variabel menggunakan teknik multicollinearity pada analisis regresi [20]. Teknik ini berperan penting untuk menentukan model persamaan regresi beserta variabel yang mempengaruhinya. Penelitian tersebut menghasilkan tingkat determinasi sebesar 0.9036.

Beberapa penelitian sebelumnya hanya fokus kepada mengidentifikasi faktor-faktor penyebab produktivitas tanaman saja. Sedangkan untuk penelitian yang lainnya sudah menerapkan metode regresi linear tetapi hanya untuk memprediksi produktivitas tanaman kelapa sawit. Belum ada penelitian yang menerapkan metode regresi linear untuk memprediksi produktivitas tanaman padi.

Oleh karena itu, berdasarkan hasil uji coba dari beberapa studi literatur terkait, pada penelitian ini digunakan penggabungan antara metode multicollinearity untuk analisis faktor

penyebab dan metode regresi linear untuk memprediksi produktivitas padi di pulau Sumatera. Kemudian, hasil uji coba dengan metode regresi linear akan dibandingkan dengan metode prediksi yang lain.

3. METODOLOGI PENELITIAN

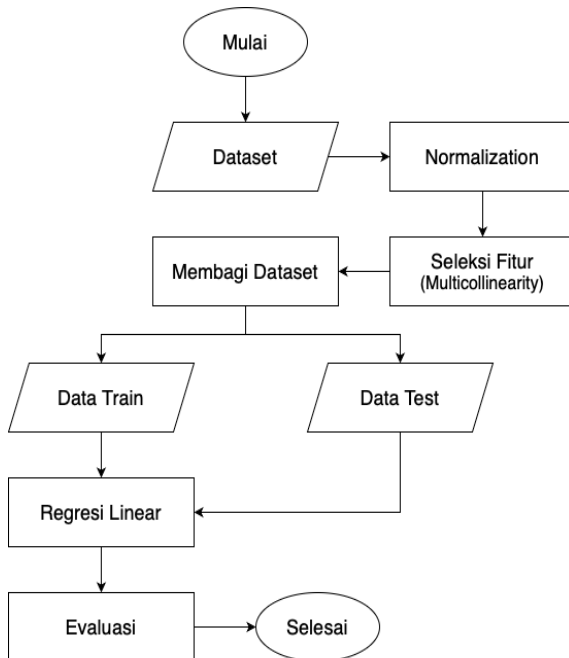
3.1. Tahapan Penelitian

Dalam penelitian ini, kami menerapkan serangkaian langkah sistematis untuk menganalisis dan memprediksi produksi padi di Pulau Sumatera menggunakan teknik predictive analytics. Secara garis besar penelitian ini terdiri dari beberapa tahapan yang dapat dilihat pada Gambar 1. Langkah pertama dalam proses ini adalah preprocessing data, yang merupakan tahap krusial untuk memastikan kualitas dan konsistensi data yang akan digunakan.

Preprocessing melibatkan beberapa aktivitas, seperti penghapusan data yang tidak relevan, penanganan data yang tidak konsisten, dan normalisasi data. Penghapusan data yang hilang dapat dilakukan dengan berbagai metode, seperti penghapusan baris atau kolom yang memiliki nilai hilang atau imputasi nilai dengan rata-rata, median, atau modus. Normalisasi data bertujuan untuk memastikan bahwa semua fitur berada dalam skala yang sama, yang penting untuk algoritma yang sensitif terhadap skala, seperti regresi linear.

Setelah preprocessing, langkah berikutnya adalah feature selection. Tujuan dari feature selection adalah untuk memilih subset fitur yang paling relevan dan signifikan untuk model prediksi. Ini membantu dalam mengurangi dimensi data, meningkatkan akurasi model, dan mengurangi overfitting.

Metode feature selection yang umum digunakan salah satunya dengan multicollinearity. Dalam konteks produksi padi, fitur yang mungkin dipertimbangkan termasuk luas lahan, curah hujan, kelembapan, dan suhu rata-rata. Pemilihan fitur yang tepat sangat penting untuk memastikan bahwa model dapat menangkap hubungan yang relevan antara variabel input dan output.



Gambar 1. Diagram Alir Tahapan Penelitian

Selanjutnya, kami menangani multicollinearity, yaitu kondisi di mana dua atau lebih fitur independen dalam dataset memiliki korelasi yang tinggi. Multicollinearity dapat menyebabkan masalah dalam model prediksi karena dapat memperbesar varians estimasi koefisien regresi, yang pada gilirannya dapat membuat model menjadi tidak stabil.

Untuk mendeteksi multicollinearity, kami menggunakan Variance Inflation Factor (VIF), yang memberikan ukuran seberapa banyak varians dari estimasi koefisien meningkat karena multicollinearity. Jika VIF dari suatu fitur lebih besar dari ambang batas tertentu biasanya 5 atau 10, fitur tersebut mungkin perlu dihapus atau digabungkan dengan fitur lain untuk mengurangi multicollinearity.

Tahap terakhir adalah prediksi menggunakan regresi linear. Regresi linear adalah metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen dan satu atau lebih variabel independen dengan cara menyesuaikan garis lurus (linear) ke data. Model regresi linear sederhana dapat dinyatakan dengan Persamaan 1.

$$y = \beta_0 + \beta_1 x + \epsilon, \quad (1)$$

Di mana y adalah variabel dependen, x adalah variabel independen, β_0 adalah intersep, β_1 adalah

koefisien regresi, dan ϵ adalah *error term*. Dalam kasus regresi linear berganda, persamaan diperluas untuk mencakup beberapa variabel independen seperti ditunjukkan pada Persamaan 2.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon, \quad (2)$$

Proses fitting model melibatkan penentuan nilai-nilai dari koefisien regresi β yang meminimalkan selisih antara nilai yang diprediksi oleh model dan nilai yang sebenarnya diamati dalam data, suatu pendekatan yang dikenal sebagai metode least squares. Setelah model dibangun, ia dapat digunakan untuk membuat prediksi tentang hasil produksi padi dengan data

Normalized Data

	Produksi	Luas Panen	Curah hujan	Kelembapan	Suhu rata-rata
0	0.012586	0.133597	0.704449	1.000000	0.359528
1	0.000000	0.111042	0.594530	0.546245	0.225933
2	0.132513	0.390448	1.000000	0.875233	0.363458
3	0.143840	0.393068	0.554475	0.862818	0.332024
4	0.055721	0.199983	0.279537	0.605214	0.412574
5	0.145440	0.413237	0.816667	0.953445	0.465619
6	0.075022	0.272066	0.727845	0.915580	0.326130
7	0.133603	0.335201	0.374410	0.878336	0.711198
8	0.152319	0.348599	0.586019	0.566108	0.418468
9	0.135503	0.267522	0.638619	0.442582	0.573674
10	0.141639	0.258593	0.452543	0.000000	1.000000
11	0.192430	0.330903	0.598040	0.792055	0.998035
12	0.205419	0.334123	0.424549	0.675978	0.394892
13	0.207751	0.326426	0.000000	0.365611	0.314342
14	0.279871	0.423918	0.557097	0.446307	0.593320
15	0.293072	0.365751	0.556573	0.638734	0.357564
16	0.427529	0.567573	0.495623	0.644941	0.416503
17	0.481605	0.631374	0.867129	0.843575	0.416503
18	0.535392	0.683085	0.406478	0.564246	0.436149
19	0.600308	0.793375	0.453673	0.651148	0.436149
20	0.642955	0.781412	0.764955	0.762880	0.416503
21	0.688638	0.815036	0.452664	0.694600	0.239686
22	0.818676	1.000000	0.430721	0.632526	0.495088
23	0.895458	0.000000	0.708846	0.663563	0.367387
24	1.000000	0.018201	0.510185	0.517070	0.349705
25	0.352697	0.382795	0.332984	0.455618	0.180747
26	0.221560	0.231634	0.462305	0.578523	0.520629
27	0.399677	0.487729	0.665967	0.440099	0.000000

Gambar 2. Hasil Normalisasi Dataset

testing. Keunggulan dari regresi linear adalah kesederhanaannya dan interpretabilitasnya, yang memungkinkan kita untuk memahami dan menjelaskan hubungan antara variabel dengan mudah.

Setelah model regresi linear dibangun dan digunakan untuk melakukan prediksi, langkah selanjutnya adalah mengevaluasi kinerjanya. Salah satu metrik evaluasi yang paling umum digunakan dalam konteks ini adalah Mean Squared Error (MSE). MSE mengukur rata-rata dari kuadrat selisih antara nilai yang diprediksi oleh model dan nilai yang sebenarnya diamati.

MSE memberikan indikasi seberapa baik model memprediksi data; nilai MSE yang lebih rendah menunjukkan bahwa model memiliki

kesalahan prediksi yang lebih kecil dan, oleh karena itu, lebih akurat.

Keuntungan utama dari MSE adalah sensitivitasnya terhadap kesalahan besar, karena kesalahan tersebut dikuadratkan sebelum dirata-ratakan, yang berarti bahwa MSE memberikan penalti lebih besar untuk kesalahan yang lebih besar. Namun, ini juga berarti bahwa MSE bisa sangat dipengaruhi oleh outlier. Oleh karena itu, penting untuk memastikan bahwa data telah dipreproses dengan baik dan outlier telah ditangani sebelum menggunakan MSE sebagai metrik evaluasi.

3.2. Dataset

Pada penelitian ini dataset yang digunakan adalah dataset produktivitas padi di seluruh provinsi di Sumatera seperti Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung. Dataset tersebut merupakan dataset *timeseries* yang bersifat *public*. Dataset produktivitas padi ini memiliki 224 data dan 7 fitur yang terdiri dari provinsi, tahun, produksi, luas panen, curah hujan, kelembapan, dan suhu rata-rata.

Dataset yang digunakan pada penelitian ini dapat dilihat pada Gambar.

Provinsi	Tahun	Produksi	Luas Panen	Curah hujan	Kelembapan	Suhu rata-rata
Aceh	2014	1820062.00	376137.00	2264.40	78.30	27.10
Aceh	2015	1956940.00	461060.00	1575.00	80.00	27.10
Aceh	2016	2180754.00	293067.00	1096.00	83.32	27.12
Aceh	2017	2478922.00	294483.00	1905.90	85.57	26.51
Aceh	2018	1751996.94	329515.78	1427.80	83.98	26.48
Aceh	2019	1714437.60	310012.46	1931.40	83.90	26.65
Aceh	2020	1861567.10	317869.41	1619.20	80.82	25.41

Gambar 3. Dataset Produktivitas Padi Di Sumatera

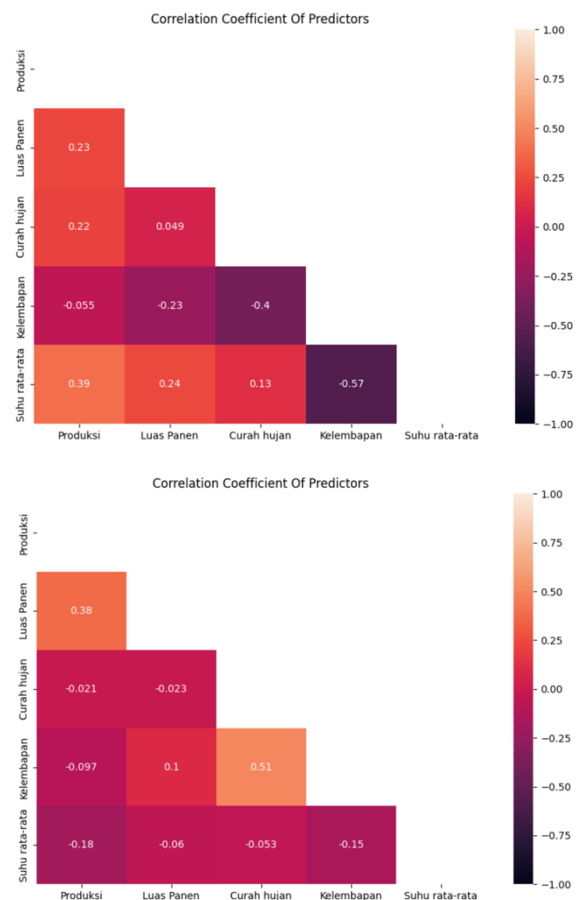
3.3. Preprocessing

Pada tahap pertama yaitu preprocessing dengan memisahkan dataset berdasarkan tiap provinsi dilanjutkan dengan normalisasi. Hasil dari normalisasi ditunjukkan pada Gambar .

Normalisasi dilakukan untuk mengubah skala menjadi 0 sampai 1 pada setiap fitur. Normalisasi bertujuan untuk mengubah skala nilai yang sama pada seluruh fitur agar menghasilkan model regresi linear yang baik.

3.4. Seleksi fitur

Tahap kedua yaitu seleksi fitur menggunakan multicollinearity, di mana bertujuan untuk mencari fitur-fitur yang memiliki korelasi yang tinggi. Maka, fitur-fitur yang memiliki korelasi rendah atau memiliki variance yang tinggi tidak dipertimbangkan sebagai variabel independen. Hubungan antar variabel menggunakan teknik multicollinearity ditunjukkan pada Gambar 4.



Gambar 2. Hubungan Antar Fitur Dengan Multicollinearity

Kemudian hubungan korelasi antar fitur dapat diukur berdasarkan nilai Variance Inflation Factor (VIF). Nilai VIF merupakan rasio atau perbandingan untuk mengukur seberapa besar kenaikan variance dari sebuah koefisien estimasi parameter regresi. Interpretasi dari nilai VIF yaitu:

1. Jika VIF = 1, maka tidak ada korelasi antar variabel independen dan tidak ada multicollinearity.

2. Jika VIF antara 1 dan 5, maka menunjukkan adanya korelasi, tetapi cukup untuk menjadi perhatian serius.
3. Jika VIF > 5, maka menunjukkan adanya multicollinearity yang dapat mempengaruhi stabilitas dan interpretasi model regresi.

Maka dari itu nilai VIF yang optimal yaitu di antara 1 dan 5. Karena jika multicollinearity yang tinggi dapat membuat estimasi koefisien regresi menjadi tidak stabil dan tidak dapat diinterpretasikan secara akurat. Solusi jika ditemukan multicollinearity yang tinggi yaitu bisa dengan menghapus variabel yang menyebabkan masalah atau menggabungkan variabel yang berkorelasi tinggi.

Contoh hasil nilai VIF pada provinsi Lampung yang dihasilkan ditunjukkan pada Gambar 5. Karena nilai VIF yang dihasilkan pada masing-masing fitur kurang dari 5 atau ideal, maka seluruh fitur tersebut memiliki korelasi yang optimal. Berdasarkan pertimbangan tersebut, maka seluruh fitur dipakai untuk membangun model regresi.

	Variable	VIF
3	Suhu rata-rata	1.260519
0	Luas Panen	1.146017
2	Kelembapan	1.125214
1	Curah hujan	1.047068
	Variable	VIF
2	Kelembapan	1.393351
1	Curah hujan	1.354493
3	Suhu rata-rata	1.026304
0	Luas Panen	1.019486

Gambar 3. Nilai VIF; (a) Provinsi Sumatera Barat, (b) Provinsi Lampung, (c) Provinsi Sumatera Utara

3.5. Regresi linear

Kemudian tahapan selanjutnya membangun model regresi, namun dilakukan pembagian

dataset menjadi data training dan data testing. Di mana data training digunakan untuk membangun model regresi linear, sedangkan data testing digunakan untuk menguji model yang sudah dibangun. Penelitian ini memiliki model regresi linear masing-masing provinsi yang terdapat dalam dataset.

Penelitian ini menggunakan regresi linear karena berdasarkan penelitian sebelumnya bekerja cukup baik jika digabungkan dengan koefisien korelasi untuk menentukan faktor-faktor yang mempengaruhi [21]. Selain itu, kelebihan yang mampu membuat model prediksi dengan jumlah sampel yang sedikit [22]. Maka dari itu, sangat tepat diimplementasikan pada penelitian ini karena dataset yang digunakan memiliki sampel terbatas untuk setiap provinsinya.

Proses fitting model dilakukan dengan memperhatikan nilai-nilai dari koefisien yang meminimalkan selisih antara nilai yang diprediksi oleh model dan nilai yang sebenarnya, seperti pada Persamaan 2. Di mana nilai-nilai koefisien tersebut merupakan nilai dari fitur-fitur yang lolos tahap seleksi fitur yaitu dengan nilai *multicollinearity* di bawah 5.

4. HASIL DAN PEMBAHASAN

Penelitian ini memiliki model regresi linear masing-masing provinsi yang terdapat dalam dataset. Setelah membangun model regresi selanjutnya dilakukan uji model regresi yang sudah dibangun. Hasil uji model regresi dilakukan dengan menggunakan teknik Mean Squared Error (MSE) seperti pada Persamaan 3.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - Y'_i)^2, \quad (3)$$

Di mana n adalah jumlah data, Y_i nilai yang diobservasi, Y'_i adalah nilai yang diprediksi. Teknik ini mengukur rata-rata dari kuadrat selisih antara nilai yang diprediksi oleh model dan nilai groundtruth. Hasil evaluasi masing-masing model tiap provinsi dapat dilihat pada **Error! Reference source not found.**

Tabel 1. Perbandingan Evaluasi Model Dengan MSE

Provinsi	Mean Squared Error (MSE)		
	Regresi Linear	SVR	RFR
Aceh	0,02	0,02	0,04

Sumatera Utara	0,04	0,40	0,03
Sumatera Barat	0,03	0,03	0,02
Riau	0,01	0,01	0,01
Jambi	0,01	0,02	0,01
Sumatera Selatan	0,02	0,02	0,02
Bengkulu	0,03	0,06	0,06
Lampung	0,03	0,04	0,02
Rata-rata	0,023	0,075	0,026

Berdasarkan hasil perbandingan hasil evaluasi MSE antara tiga metode yaitu Regresi Linear, Support Vector Regression (SVR), Random Forest Regression (RFR) menghasilkan nilai rata-rata MSE terendah sebesar 0,023 dengan metode Regresi Linear. Namun, nilai rata-rata MSE dari Regresi Linear tidak jauh signifikan jika dibandingkan dengan RFR.

Dari hasil uji coba menunjukkan metode regresi linear lebih baik jika diterapkan untuk dataset yang jumlah sampelnya terbatas. Berbanding terbalik dengan SVR dan RFR. Namun, hasilnya tidak terlalu signifikan.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan menjadi beberapa poin:

1. Dataset produksi padi di Sumatera yang digunakan pada penelitian ini sudah memiliki nilai korelasi yang cukup baik karena memiliki VIF di bawah 5 atau optimal.
2. Metode Regresi Linear jauh lebih baik jika dibandingkan dengan metode SVR, namun tidak terlalu signifikan jika dibandingkan dengan metode RFR.
3. Teknik multicollinearity sudah cukup representatif dalam seleksi fitur dengan cara mencari korelasi antar fitur.
4. Penggabungan antara metode multicollinearity untuk analisis variabel dan metode regresi linear untuk prediksi sangat baik digunakan untuk dataset produktivitas padi.

Kemudian untuk saran yang bisa dikembangkan dalam penelitian ini yaitu:

1. Mengembangkan model non-linear, mengingat bahwa hubungan antar variabel tidak sepenuhnya linear.
2. Memasukkan data cuaca dan iklim pada dataset supaya lebih komprehensif sehingga mampu meningkatkan prediksi produktivitas padi.
3. Menguji model pada berbagai lokasi geografis dan musim tanam yang berbeda dapat membantu mengevaluasi model.
4. Pengembangan alat prediksi berbasis web atau aplikasi dapat membantu petani dan stakeholder dalam membuat keputusan yang lebih baik.

6. UCAPAN TERIMA KASIH

Keberhasilan penulis dalam menyelesaikan penelitian ini tidak lepas dari beberapa pihak yang telah membantu. Penulis mengucapkan terima kasih kepada Universitas Sebelas Maret yang telah memberikan dorongan serta kesempatan sehingga penelitian ini berjalan maksimal.

Selain itu, penulis juga mengucapkan terima kasih kepada Politeknik Elektronika Negeri Surabaya yang telah membantu dan berkolaborasi dalam penelitian ini. Serta ucapan terima kasih disampaikan kepada semua rekan yang terlibat dalam melaksanakan kegiatan penelitian ini.

Dengan penelitian ini diharapkan mampu memberikan manfaat bagi *stakeholder* di bidang pertanian. Di mana setiap pengambilan keputusan berbasis data, sehingga dapat membuat keputusan yang lebih efektif dan efisien. Kemudian di tingkat pemerintah, model prediksi ini bisa digunakan untuk merencanakan distribusi sumber daya dan program dukungan lainnya untuk petani di Sumatera.

DAFTAR PUSTAKA:

- [1] B. P. S. Indonesia, "Ringkasan Eksekutif Luas Panen dan Produksi Padi di Indonesia 2023 (Angka Tetap)." Accessed: Dec. 18, 2024. [Online]. Available: <https://www.bps.go.id/id/publication/2024/05/06/69834d72f7ef1c32eee5c4b6/ringkasan-eksekutif-luas-panen-dan-produksi-padi-di-indonesia-2023--angka-tetap-.html>



- [2] S. Sulaminingsih, E. Silamat, A. Ruruh, M. Syaiful, A. Ninasari, and M. Ar, "DAMPAK PERUBAHAN IKLIM TERHADAP PENINGKATAN DAN PENURUNAN PRODUKTIVITAS TANAMAN PANGAN," *J. Rev. Pendidik. Dan Pengajaran JRPP*, vol. 7, no. 3, pp. 10189–10195, Jul. 2024, doi: 10.31004/jrpp.v7i3.31609.
- [3] S. Maesaroh and K. K, "Sistem Prediksi Produktifitas Pertanian Padi Menggunakan Data Mining," *Energy J. Ilm. Ilmu-Ilmu Tek.*, vol. 7, no. 2, Art. no. 2, Nov. 2017.
- [4] M. F. Anggarda, I. Kustiawan, D. R. Nurjanah, and N. F. A. Hakim, "Pengembangan Sistem Prediksi Waktu Penyiraman Optimal pada Perkebunan: Pendekatan Machine Learning untuk Peningkatan Produktivitas Pertanian," *J. Budid. Pertan.*, vol. 19, no. 2, Art. no. 2, Dec. 2023, doi: 10.30598/jbdp.2023.19.2.124.
- [5] N. H. Darlan, I. Pradiko, and H. H. Siregar, "PREDIKSI DAN ANTISIPASI KEJADIAN CUACA EKSTRIM DAN DAMPAKNYA TERHADAP PRODUKSI TANAMAN KELAPA SAWIT".
- [6] *The State of Food and Agriculture 2022*. FAO, 2022. doi: 10.4060/cb9479en.
- [7] K. Prayoga, "DAMPAK PENETRASI TEKNOLOGI INFORMASI DALAM TRANSFORMASI SISTEM PENYULUHAN PERTANIAN DI INDONESIA," *JSEP J. Soc. Agric. Econ.*, vol. 11, no. 1, Art. no. 1, Jun. 2018, doi: 10.19184/jsep.v11i1.5663.
- [8] D. Marpaung, S. Sumarno, and I. Gunawan, "Prediksi Produktivitas Kelapa Sawit di PTPN IV dengan Algoritma Backpropagation," *KLIK Kaji. Ilm. Inform. Dan Komput.*, vol. 1, no. 2, Art. no. 2, Oct. 2020, doi: 10.30865/klik.v1i2.48.
- [9] A. Wanto, "PREDIKSI PRODUKTIVITAS JAGUNG DI INDONESIA SEBAGAI UPAYA ANTISIPASI IMPOR MENGGUNAKAN JARINGAN SARAF TIRUAN BACKPROPAGATION," *SINTECH Sci. Inf. Technol. J.*, vol. 2, no. 1, Art. no. 1, Apr. 2019, doi: 10.31598/sintechjournal.v2i1.355.
- [10] J. Campoy, I. Campos, C. Plaza, M. Calera, V. Bodas, and A. Calera, "Estimation of harvest index in wheat crops using a remote sensing-based approach," *Field Crops Res.*, vol. 256, p. 107910, Oct. 2020, doi: 10.1016/j.fcr.2020.107910.
- [11] J. Xu, J. Meng, and L. J. Quackenbush, "Use of remote sensing to predict the optimal harvest date of corn," *Field Crops Res.*, vol. 236, pp. 1–13, Apr. 2019, doi: 10.1016/j.fcr.2019.03.003.
- [12] D. Ariani, Y. Prasetyo, and B. Sasmito, "ESTIMASI TINGKAT PRODUKTIVITAS PADI BERDASARKAN ALGORITMA NDVI, EVI DAN SAVI MENGGUNAKAN CITRA SENTINEL-2 MULTITEMPORAL (Studi Kasus: Kabupaten Pekalongan, Jawa Tengah)," *J. Geod. Undip*, vol. 9, no. 1, Art. no. 1, Dec. 2019, doi: 10.14710/jgundip.2020.26165.
- [13] M. Kurniasari and P. G. Ariastita, "Faktor - Faktor yang Mempengaruhi Alih Fungsi Lahan Pertanian Sebagai Upaya Prediksi Perkembangan Lahan Pertanian di Kabupaten Lamongan," *J. Tek. ITS*, vol. 3, no. 2, pp. C119–C124, Sep. 2014, doi: 10.12962/j23373539.v3i2.7237.
- [14] J. H. Galitan, F. Duko, and F. Hatim, "Analisis Produksi Padi Di Indonesia: Indonesia," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 2, Art. no. 2, Apr. 2024.
- [15] I. Permatasari, Gusriati, and H. Gusvita, "ANALISIS FAKTOR-FAKTOR PRODUKSI PADI SAWAH DAN HUBUNGANNYA DENGAN PENDAPATAN PETANI.," *UNES J. Mhs. Pertan.*, vol. 2, no. 1, Art. no. 1, Apr. 2018.
- [16] D. Yanti, Martanto, and A. Bahtiar, "Prediksi Hasil Panen Padi Tahun 2023 menggunakan Metode Regresi Linier di Kabupaten Indramayu," *J. Inform. Terpadu*, vol. 9, no. 1, Art. no. 1, Mar. 2023, doi: 10.54914/jit.v9i1.657.
- [17] E. Triyanto, H. Sismoro, and A. D. Laksito, "IMPLEMENTASI ALGORITMA REGRESI LINEAR BERGANDA UNTUK MEMPREDIKSI PRODUKSI PADI DI KABUPATEN BANTUL," *Rabit J. Teknol. Dan Sist. Inf. Univrab*, vol. 4, no. 2, pp. 66–75, Jul. 2019, doi: 10.36341/rabit.v4i2.666.
- [18] O. Muharromah, N. Suarna, and W. Prihartono, "IMPLEMENTASI ALGORITMA REGRESI LINEAR BERGANDA UNTUK PREDIKSI PRODUKSI PADI DI KABUPATEN CIREBON," *JATI J. Mhs. Tek. Inform.*, vol. 7, no. 6, Art. no. 6, 2023, doi: 10.36040/jati.v7i6.8206.
- [19] N. Aswan, Y. Fadhillah, M. N. H. Siregar, N. R. Puspita, T. Mahaji, and A. U. Harahap,



- "Analisis Regresi Linier dalam Estimasi Produktivitas Kelapa Sawit (*Elaeis guineensis*) Di Kabupaten Tapanuli Selatan," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 1, Art. no. 1, Feb. 2024, doi: 10.31004/innovative.v4i1.8883.
- [20] M. Sriningsih, D. Hatidja, and J. D. Prang, "PENANGANAN MULTIKOLINEARITAS DENGAN MENGGUNAKAN ANALISIS REGRESI KOMPONEN UTAMA PADA KASUS IMPOR BERAS DI PROVINSI SULUT," *J. Ilm. Sains*, pp. 18–24, Jul. 2018, doi: 10.35799/jis.18.1.2018.19396.
- [21] Q. Zhang, "Housing Price Prediction Based on Multiple Linear Regression," *Sci. Program.*, vol. 2021, no. 1, p. 7678931, 2021, doi: 10.1155/2021/7678931.
- [22] A. C. P. Fernandes, A. R. Fonseca, F. A. L. Pacheco, and L. F. Sanches Fernandes, "Water quality predictions through linear regression - A brute force algorithm approach," *MethodsX*, vol. 10, p. 102153, Jan. 2023, doi: 10.1016/j.mex.2023.102153.