

## OPTIMASI KINERJA *LINEAR REGRESSION*, *RANDOM FOREST REGRESSION* DAN *MULTILAYER PERCEPTRON* PADA PREDIKSI HASIL PANEN

Evita Fitri <sup>1\*</sup>; Siti Nurhasanah Nugraha <sup>2</sup>

Sistem Informasi <sup>1,2</sup>  
Universitas Nusa Mandiri <sup>1,2</sup>  
<https://www.nusamandiri.ac.id/> <sup>1,2</sup>  
evita.etv@nusamandiri.ac.id <sup>1\*</sup>, siti.nhg@nusamandiri.ac.id <sup>2</sup>



**Abstract**—Rice yield prediction is a significant challenge in the context of climate uncertainty and farmland variation. Erratic weather factors, along with land differences, make this prediction more complex. This research aims to address these issues using a machine learning approach. The method used involves three machine learning models namely Linear regression, Random Forest Regression, and ANN with MultiLayer Perceptron algorithm as well as the evaluation matrix RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error). This research focuses on testing the accuracy of the three models in the face of uncertain seasonal conditions and variations in agricultural land. The results showed that the MultiLayer Perceptron prediction model gave the best results with an error value of 0.094. The random forest regression method ranks second with an error value of 0.510, followed by Linear regression with an error value of 0.281. The importance of outlier testing in the model development process can be seen from the significant improvement in the performance of the MultiLayer Perceptron model. This research contributes to the development of a more reliable and dependable rice yield prediction system, especially in the midst of uncertain climatic conditions. Machine learning models, particularly MultiLayer Perceptron, can be an effective solution to increase agricultural productivity and reduce risks associated with weather changes and land variations.

**Keywords:** linear regression, multilayer perceptron, random forest regression, regression model, rice prediction.

**Abstrak**—Prediksi hasil panen padi menjadi tantangan yang signifikan dalam konteks ketidakpastian iklim dan variasi lahan pertanian. Faktor cuaca yang tidak menentu, bersamaan dengan perbedaan lahan, membuat prediksi ini semakin kompleks. Penelitian ini bertujuan untuk mengatasi permasalahan tersebut dengan menggunakan pendekatan *machine learning*. Metode yang digunakan melibatkan tiga model *machine learning* yaitu *Linear regression*, *Random Forest Regression*, dan ANN dengan algoritma *MultiLayer Perceptron* serta dengan matrix evaluasi RMSE (*Root Mean Squared Error*), MAE (*Mean Absolute Error*) dan MAPE (*Mean Absolute Percentage Error*). Penelitian ini fokus pada pengujian akurasi ketiga model tersebut dalam menghadapi kondisi musim yang tidak menentu dan variasi lahan pertanian. Hasil penelitian menunjukkan bahwa model prediksi *MultiLayer Perceptron* memberikan hasil terbaik dengan nilai *error* sebesar 0.094. Metode *Random Forest Regression* menempati urutan kedua dengan nilai *error* 0.510, diikuti oleh *Linear regression* dengan nilai *error* 0.281. Pentingnya pengujian *outlier* dalam proses pengembangan model terlihat dari peningkatan signifikan dalam performa model *MultiLayer Perceptron*. Penelitian ini memberikan kontribusi pada pengembangan sistem prediksi hasil panen padi yang lebih handal dan dapat diandalkan, terutama di tengah ketidakpastian kondisi iklim. Model *machine learning*, khususnya *MultiLayer Perceptron*, dapat menjadi solusi efektif untuk meningkatkan produktivitas pertanian dan mengurangi risiko yang terkait dengan perubahan cuaca dan variasi lahan.

**Kata kunci:** linear regression, multilayer perceptron, random forest regression, model regresi, prediksi padi.

### PENDAHULUAN

Pertanian merupakan sektor vital dalam pemenuhan kebutuhan pangan global, dan produksi

padi sebagai salah satu komoditas utama menjadi fokus utama dalam rangka menjaga ketahanan pangan. Namun, fluktuasi iklim dan perubahan cuaca yang semakin tidak terduga dapat berdampak

signifikan terhadap produktivitas pertanian, khususnya pada proses panen padi. Kondisi iklim yang tidak menentu, seperti perubahan pola hujan dan suhu ekstrem, telah memberikan dampak yang signifikan terhadap produktivitas pertanian. Perubahan iklim yang berdampak pada pola cuaca yang tidak terduga sering kali menyebabkan kerugian yang signifikan dalam sektor pertanian. Peningkatan suhu, perubahan pola hujan, dan kejadian cuaca ekstrem juga dapat mengganggu siklus pertumbuhan tanaman dan mengurangi hasil panen secara drastis, hal ini sejalan dengan penelitian (Nuraisah & Kusumo, 2019). Selain itu, variasi lahan pertanian yang mencakup berbagai tipe tanah dan topografi juga turut mempengaruhi suatu hasil panen (Herlina & Prasetyorini, 2020). Oleh karena itu, pengembangan model prediksi yang dapat mengatasi permasalahan tersebut menjadi esensial dalam upaya meningkatkan produktivitas dan ketahanan pangan serta menjadi sangat penting untuk membantu petani dan pemangku kepentingan terkait dalam mengambil keputusan yang tepat waktu guna mengoptimalkan hasil panen.

Dalam era transformasi digital, penerapan teknologi kecerdasan buatan, khususnya metode *machine learning*, telah menjadi suatu terobosan yang menjanjikan dalam berbagai bidang, salah satunya pada bidang pertanian (Sumadi et al., 2022). *Machine learning* mampu memanfaatkan data terkait iklim, tanah dan faktor-faktor pertanian lainnya untuk memberikan prediksi yang lebih akurat dan terperinci, serta memberikan keuntungan signifikan bagi para petani.

Pada prediksi hasil panen tanaman pangan yang dilakukan oleh (Satria et al., 2023), penelitian tersebut menganalisis pentingnya sektor pertanian di Indonesia, khususnya di pulau Sumatera, yang berkontribusi pada peningkatan produksi pangan, industri, dan ekspor. Meskipun pulau Sumatera produktif, pertanian di daerah tersebut rentan terhadap perubahan iklim global. Adapun metode yang digunakan dalam penelitian (Satria et al., 2023) menggunakan beberapa algoritma, yaitu *Random Forest* (RF), *Decision Tree* (DT), *Gradient Boosting* (GB), *Extra Tree* (ET), *Support Vector Machine* (SVM), dan *Artificial Neural Network* (ANN). Hasil penelitian menunjukkan bahwa nilai koefisien  $R^2$  tertinggi diperoleh oleh model *Extra Tree*, yang lebih akurat dalam memprediksi hasil pertanian dibandingkan dengan model lainnya dengan score  $R^2$  0,968.

Selanjutnya perbedaan model prediksi yang dilakukan pada penelitian (Herwanto et al., 2019) menggunakan metode *Linear regression*, penelitian ini melibatkan pengumpulan data dari petani di Lamongan melalui angket, dilanjutkan dengan *pre-*

*processing* data, penerapan *Linear regression*, dan validasi hasil. Hasil pengujian akurasi dengan RMSE (*Root Mean Squared Error*) menunjukkan rata-rata akurasi sebesar 0,432, menandakan bahwa model prakiraan memiliki tingkat akurasi yang baik. Model *multiple Linear regression* juga dinyatakan cocok dengan tingkat keandalan sebesar 94,51%.

Berbeda dengan penelitian (Putra & Walmi, 2020) terkait penerapan prediksi produksi padi menggunakan *Artificial Neural Network* (ANN) algoritma *backpropagation*, permasalahan utama yang diatasi dalam artikel ini adalah bagaimana menentukan jumlah *neuron* dan *hidden layer* yang optimal dalam arsitektur ANN untuk memprediksi produksi padi dengan tingkat akurasi yang tinggi. Proses ini menjadi krusial karena penentuan struktur optimal ANN dapat mempengaruhi kemampuan sistem dalam melakukan prediksi produksi padi dengan tepat. Namun hasil dari penelitian tersebut menunjukkan bahwa setelah diuji pada 19 daerah di Sumatera Barat, sistem prediksi produksi padi yang menggunakan arsitektur ANN dengan algoritma *backpropagation* mencapai tingkat akurasi sebesar 88,14%. Tingkat *error* yang diperoleh juga relatif rendah, yaitu sebesar 11,86%.

Adapun pada penelitian (Wardhani et al., 2022) yang memprediksi hasil panen menggunakan algoritma deep learning dengan menguji dataset sebanyak 380 baris yang dikumpulkan melalui penyebaran angket. Adapun evaluasi yang dilakukan yaitu dengan melihat nilai loss MSE (*Mean Squared Error*), MAE (*Mean Absolute Error*), dan MAPE (*Mean Absolute Percentage Error*) berturut-turut sebesar 2939977.418, 301.788, dan 83.798. Namun dapat dilihat bahwa hasil nilai MAPE cukup tinggi, dimana score MAPE mengukur persentase rata-rata kesalahan prediksi, dan nilai MAPE yang dihasilkan oleh penelitian tersebut sebesar 83.798% sebenarnya menunjukkan tingkat kesalahan yang relatif tinggi dalam prediksi hasil panen padi yang seharusnya semakin nilai MAPE mendekati nol, semakin baik performa modelnya.

Pada penelitian lainnya (Evita, 2021) melakukan prediksi hasil panen menggunakan metode ANN dengan algoritma *backpropagation* sebagai solusi untuk merancang arsitektur ANN yang dapat memberikan prediksi produksi jagung dengan tingkat akurasi tinggi. Masalah utama yang dihadapi adalah menentukan jumlah *neuron* dan *hidden layer* yang optimal untuk meningkatkan akurasi prediksi dengan memperhatikan parameter seperti *epoch*, *momentum*, *learning rate*, dan *hidden layer* hasil akurasi mencapai 88,14%.

Dapat dilihat bahwa dari pengujian beberapa model prediksi yang dilakukan oleh para peneliti, menghasilkan model terbaik diantaranya *Extra*

Tree dengan score  $R^2$  0,968 pada penelitian (Satria et al., 2023), *Linear regression* dengan RMSE 0,432 dan akurasi 94,51% pada penelitian (Herwanto et al., 2019), lalu pada penelitian (Putra & Walmi, 2020) dan (Evita, 2021) dengan model terbaiknya yaitu ANN dengan algoritma backpropagation sebesar tingkat akurasi sebesar 88,14%. Serta nilai *error* sebesar 11,86%, dan terakhir model prediksi *deep learning* dengan nilai loss MSE, MAE, dan MAPE sebesar 2939977.418, 301.788, dan 83.798 pada penelitian (Wardhani et al., 2022). Berkaitan dengan kelima penelitian tersebut, dua diantaranya terbukti bahwa model ANN dapat menjadi model prediksi yang baik untuk pengujian prediksi hasil panen.

Adapun penelitian ini bertujuan untuk menguji potensi dan keunggulan metode *machine learning* dalam meramalkan hasil panen padi, dengan tujuan meningkatkan efisiensi produksi pertanian dan menghadapi tantangan yang muncul salah satunya akibat perubahan iklim. Penelitian ini juga relevan dalam upaya mewujudkan ketahanan pangan global. Pada penelitian ini, dilakukan pengujian terhadap dataset padi di suatu wilayah dengan beberapa feature atau variable, terdapat variabel dependen dan beberapa variabel bebas, pengujian pada penelitian ini berfokus pada model prediksi ANN dengan algoritma *MultiLayer Perceptron* disertai dengan proses penanganan *outliers*, juga menguji model lainnya sebagai pembandingan, diantaranya *Linear Regression* dengan tipe *Multiple Linear regression* dan *Random Forest Regression*. *Random Forest Regression* salah satu teknik dalam *machine learning* yang menerapkan konsep *ensemble learning*, terutama dalam konteks tugas regresi. Dalam *ensemble learning*, hasil dari beberapa model digabungkan untuk meningkatkan kinerja dan akurasi prediksi dibandingkan dengan menggunakan satu model tunggal serta melibatkan pengujian menggunakan konsep supervised dalam konstruksi kelas *classifier*, algoritma ini menggabungkan prediksi berdasarkan *Multiple Decision Tree* (Fitri, 2023).

*Multiple Linear Regression* merupakan penganalisisan untuk memproyeksikan bagaimana perubahan dalam satu variabel dapat dipengaruhi oleh perubahan dalam variabel lain disebut juga sebagai regresi berganda. Istilah ini merujuk pada situasi di mana lebih dari satu variabel bebas digunakan sebagai prediktor untuk memahami perubahan nilai variabel tertentu (Alita et al., 2021). Sedangkan *MultiLayer Perceptron* adalah jaringan saraf yang melatih menggunakan pembelajaran propagas (Rasna & Matdoan, 2022). Pada *MultiLayer Perceptron*, jaringan *feedforward* dikategorikan sebagai *perceptron multilayer*. Lapisan input, lapisan tersembunyi, dan lapisan

output adalah tiga lapisan node yang membentuk model *MultiLayer Perceptron*. Pemilihan elemen *MultiLayer Perceptron* harus dikaji dengan cermat karena memiliki dampak yang signifikan terhadap kinerja jaringan. Metode pembelajaran, fungsi aktivasi, dan desain jaringan merupakan komponen-komponen dari *MultiLayer Perceptron* (Wibawa et al., 2020).

Pada penelitian ini, pemilihan metrik evaluasi dari pengujian model prediksi yaitu menggunakan RMSE (*Root Mean Squared Error*) dan MAE (*Mean Absolute Error*) sebagai tolak ukur dari seberapa besar kesalahan prediksi dan seberapa seberapa baik model menyesuaikan dengan variasi data sebenarnya, adapun nilai MAPE (*Mean Absolute Percentage Error*). digunakan pada evaluasi ini untuk mengukur rata-rata persentase kesalahan prediksi, memberikan gambaran relatif kesalahan dalam bentuk persentase. RMSE (*Root Mean Squared Error*) merupakan metrik yang mengukur akurasi prediksi dengan menghitung akar kuadrat dari rata-rata kuadrat selisih antara nilai prediksi dan nilai sebenarnya. RMSE menjadi salah satu aturan penilaian kuadrat yang juga mengukur ukuran rata-rata kesalahan, nilai RMSE yang lebih rendah menunjukkan akurasi yang lebih baik (Milniadi & Adiwijaya, 2023). MAE sendiri merupakan metrik untuk mengevaluasi perhitungan antara nilai sebenarnya dengan nilai prediksi (Zulvian, Prihandani, & Ridha, 2021) sedangkan metrik MAPE ini mengukur rata-rata dalam bentuk persentase untuk mengukur kesalahan prediksi (Dardanella, Hidayat, Santosa, & Siskandar, 2022).

## BAHAN DAN METODE

Pada penelitian ini, Gambar 1 merupakan rancangan penelitian secara garis besar yang dilakukan, berikut *concept map* penelitian yang merupakan detail alur proses yang dilakukan pada penelitian ini. Adapun penelitian dimulai dari tahapan *Pra-Pre-processing Data*, dilanjutkan dengan *Pre-processing* dan diakhiri dengan *Pemodelan* serta *Evaluasi*.

### Pra Pre-processing

Tahapan *Pra-preprocessing* yang melibatkan identifikasi masalah dalam prediksi hasil panen yang berfungsi sebagai langkah awal untuk merinci dan memahami secara mendalam aspek-aspek kritis yang relevan dengan tujuan penelitian. Setelah itu, dilanjutkan dengan melakukan studi literatur untuk menggali pemahaman mendalam terkait dengan kerangka kerja konseptual dan temuan-temuan sebelumnya yang terkait dengan prediksi hasil panen. Selanjutnya dilakukan

pengumpulan dataset sehingga keseluruhan rangkaian tahapan ini menciptakan dasar yang kokoh untuk mengeksplorasi lebih lanjut dan mengembangkan model prediksi hasil panen dengan landasan pengetahuan yang kuat dari identifikasi masalah hingga literatur dan data.

**Pre-processing Data**

Tahapan *Pre-processing* dalam rangkaian penelitian ini adalah proses pra-pemrosesan data yang melibatkan serangkaian langkah seperti filter kolom, penanganan nilai yang hilang atau *missing value*, normalisasi, numeric *outliers*, dan pembagian data menjadi partisi yang relevan. Tahapan-tahapan ini bertujuan untuk memastikan kebersihan dan konsistensi data, serta mempersiapkan dataset dengan baik sebelum dilibatkan dalam analisis lebih lanjut.

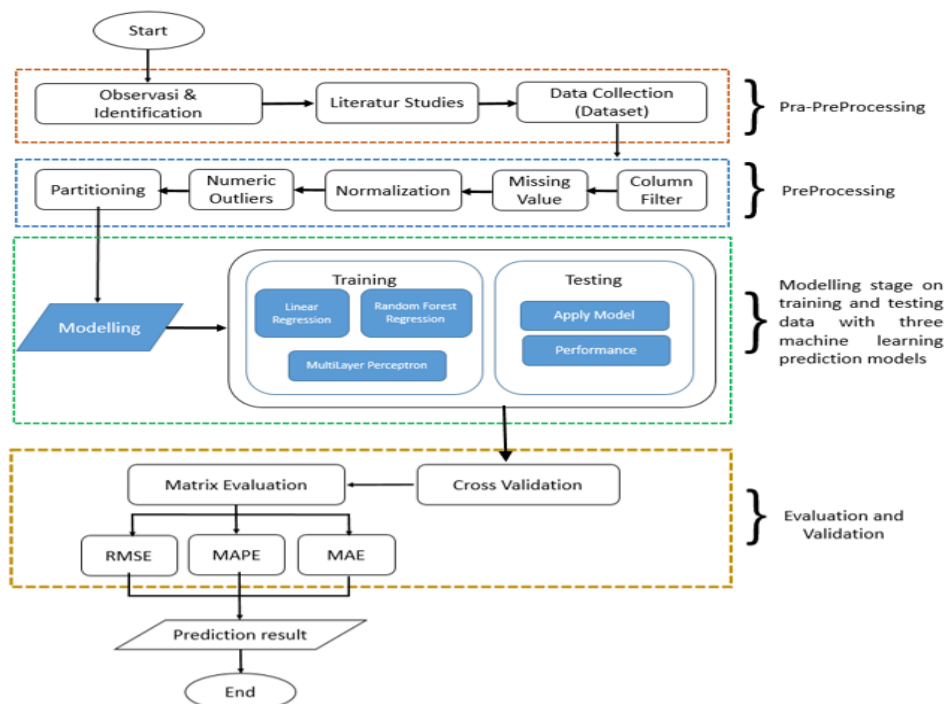
**Pemodelan Data**

Tahapan Modelling dalam rangkaian penelitian ini menerapkan tiga model yang berbeda, yaitu model regresi linear, model regresi hutan acak (*random forest*), dan model jaringan saraf tiruan *MultiLayer Perceptron*. Proses ini bertujuan untuk menganalisis dan membandingkan performa ketiga model dalam meramalkan hasil panen berdasarkan data latih. Setelah melatih model dengan data pelatihan, langkah berikutnya adalah menguji model tersebut menggunakan data uji. Kinerja

setiap model dievaluasi secara cermat untuk menilai sejauh mana kemampuannya dalam meramalkan hasil panen dengan akurat. Dengan pendekatan ini, penelitian dapat mengidentifikasi model yang paling sesuai untuk meramalkan hasil panen berdasarkan karakteristik dataset yang telah diolah.

**Evaluasi dan Validasi Data**

Proses akhir dalam rangkaian penelitian ini melibatkan tahap evaluasi dan validasi dengan menggunakan *cross-validation*. Pada tahap ini, dilakukan analisis terhadap performa ketiga model prediksi dengan fokus pada penilaian menggunakan metrik evaluasi seperti RMSE, MAE dan MAPE. Skor nilai RMSE memberikan gambaran tentang seberapa baik model dapat meminimalkan kesalahan prediksi, sedangkan MAE memberikan ukuran kesalahan rata-rata secara absolut dan nilai MAPE memberikan perspektif terhadap akurasi model dalam konteks presentasi kesalahan relatif. Selain itu, *Cross-validation* digunakan sebagai pendekatan validasi yang memastikan ketepatan hasil model dengan cara membagi data menjadi subset yang saling bergantian untuk proses pelatihan dan pengujian. Hal ini membantu menghindari bias yang mungkin timbul akibat pengujian pada satu set data tertentu.



Sumber: (Fitri & Nugraha, 2024)

Gambar 1. *Concept Map* Penelitian

**Dataset Penelitian**

Adapun dataset yang digunakan pada penelitian ini diambil dari Kaggle dengan nama dataset “Dataset Tanaman Padi Sumatera, Indonesia” yang merupakan data histori produktifitas hasil panen padi dari 8 provinsi di pulau Sumatera yang diperoleh dari Badan Pusat Statistik Indonesia, dan data perubahan cuaca yang diperoleh dari BMKG, dengan detail seperti yang ditampilkan pada tabel 1.

Tabel 1. Dataset Penelitian

No	Keterangan	Tahun	Atribut	Jumlah Instance
1	Produktifitas Hasil Panen Padi Pulau Sumatera	1993 - 2020	Provinsi, Tahun, Hasil Produksi Panen, Luas Panen	224
2	Data Perubahan Cuaca	1993 - 2020	Curah Hujan, Kelembapan, Suhu Rata-Rata	224

Sumber : (Kaggle, 2022)

Pada Tabel 1 histori data panen yang digunakan meliputi 8 provinsi diantaranya Sumatera Utara, Nanggroe Aceh Darussalam (NAD), Sumatera Selatan, Riau, Jambi, Bengkulu, Lampung dan Sumatera Barat.

**HASIL DAN PEMBAHASAN**

Penelitian ini dilakukan menggunakan histori data hasil panen yang berjumlah 224 instance, meliputi beberapa variabel independen yaitu wilayah, tahun, luas lahan panen, curah hujan, kelembapan dan suhu rata-rata, sedangkan variabel dependen yang ditetapkan pada uji coba penelitian ini yaitu jumlah hasil produksi panen pada setiap tahunnya. Adapun detail terkait dataset yang digunakan pada penelitian ini ditampilkan pada tabel 2.

Tabel 2. Sample Dataset

Provin si	Ta hu n	Prod uksi	Luas Pane n	Cura h huja n	Kele mbap an	Suhu rata-rata
Aceh	19 93	1329 536.00	3235 89.00	1627.00	82.00	26.06.00
Sumate ra Utara	19 93	2918 152.00	7545 69.00	2549.40	83.80	26.45.00
Sumate ra Barat	19 93	1806 424.00	3944 12.00	5110.00	89.77	24.70
Riau	19 93	4362 97.00	1461 33.00	2738.40	86.37	26.00.00

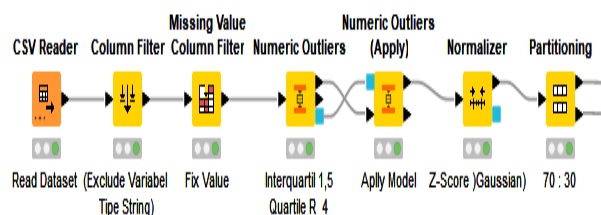
Provin si	Ta hu n	Prod uksi	Luas Pane n	Cura h huja n	Kele mbap an	Suhu rata-rata
Jambi	19 93	6075 29.00	1994 31.00	1954.00	83.81	26.29.00
Sumate ra Selatan	19 93	1409 559.00	4398 95.00	2815.00	82.78	26.38.00
Bengku lu	19 93	3567 09.00	1098 07.00	4150.00	85.50	25.95
Lampu ng	19 93	1646 900.00	4330 78.00	2306.70	84.82	26.41.00

Sumber : (Kaggle, 2022)

Tabel 2 menunjukkan sample dataset dari masing-masing provinsi yang belum dilakukan pemrosesan pembersihan data. Selanjutnya dilakukan proses *pre-processing* data untuk memastikan kevalidan, kebersihan, dan konsistensi data sebelum dilibatkan dalam analisis lebih lanjut.

**Pre-processing Dataset**

Tahapan *pre-processing* yang melibatkan beberapa rangkaian dapat dilihat pada visualisasi Gambar 2.



Sumber: (Fitri & Nugraha, 2024)

Gambar 2. Visualisasi *Pre-processing*

Pada Gambar 2 terkait *pre-processing* data setelah pembacaan dataset, proses ini dimulai dengan *Column Filter*, dengan melakukan seleksi kolom yang relevan dan signifikan untuk keperluan analisis prediksi hasil panen yaitu dengan tidak mengikut sertakan variabel dengan tipe string yaitu provinsi sehingga atribut yang digunakan hingga akhir penelitian yaitu 6 fitur.

Tahapan selanjutnya mencakup penanganan nilai yang hilang (*missing value*), di mana upaya ini dilakukan untuk mengidentifikasi dan mengatasi nilai yang tidak lengkap atau absen dalam dataset. Langkah ini esensial untuk memastikan integritas data yang digunakan dalam perhitungan dan model prediksi. Pada proses *missing value* diimplementasikan dengan nilai threshold 90%, sehingga set data, variabel atau kolom tertentu dapat diterima jika jumlah *missing value* dalam variabel tersebut tidak melebihi 90% dari total observasi atau entri dalam variabel tersebut.

Proses penanganan *outliers* juga dilakukan untuk menjaga nilai akurasi dalam analisis, meningkatkan kualitas modelling, mengurangi bias, mendukung interpretasi data serta menjaga kesalahan generalisasi pada uji coba suatu model. Adapun metode pengukuran pada proses penanganan *outliers* menggunakan *Interquartile* (IQR) atau ambang batas atas dan bawah sebesar 1,5 dan nilai *Quartil R\_4*, Dapat dilihat pada Tabel 3 bahwa terdapat jumlah *outliers* dari masing-masing variabel.

Tabel 3. Analisa Jumlah *Outliers* Pada Setiap Variabel

<i>Outlier Column</i>	<i>Member Count</i>	<i>Outlier Count</i>	<i>Lower Bound</i>	<i>Upper Bound</i>
Tahun	224	0	1,978	2,034
Hasil Produksi Panen	224	0	2,284,0 83.5	5,259,0 64.5
Luas Panen	224	0	406,102 .0	1,066,6 42.0
Curah Hujan	224	3	-321.85	5,056.1 5
Kelembapan	224	14	71,25	91.65
Suhu Rata-rata	224	28	24.625	28.745

Sumber: (Fitri & Nugraha, 2024)

Tabel 3 merupakan identifikasi *outlier* yang hasilnya ditemukan data *outliers* pada variabel curah hujan, kelembapan serta suhu rata-rata, adapun hasil dari pengujian data dalam penanganan *outliers* dapat terlihat pada tabel 4.

Tabel 4. Hasil Proses Penanganan *Numeric Outliers*

<i>Outlier Column</i>	<i>Member Count</i>	<i>Outlier Count</i>	<i>Lower Bound</i>	<i>Upper Bound</i>
Tahun	187	0	1,978	2,034
Hasil Produksi Panen	187	0	2,284,0 83.5	5,259,0 64.5
Luas Panen	187	0	406,102 .0	1,066,6 42.0
Curah Hujan	187	0	-321.85	5,056.1 5
Kelembapan	187	0	71,25	91.65
Suhu Rata-rata	187	0	24.625	28.745

Sumber: (Fitri & Nugraha, 2024)

Pada Tabel 4 terlihat bahwa setelah dilakukannya proses penanganan *outliers*, sudah tidak ditemukannya *outliers* dari masing-masing variabel sehingga data dapat dilanjutkan dengan proses normalisasi. Tahap normalisasi melibatkan untuk menyelaraskan skala dan rentang nilai pada dataset. Proses ini mendukung konsistensi interpretasi dan perbandingan antar variabel, memastikan bahwa pengaruh dari masing-masing

variabel tidak terdistorsi akibat perbedaan skala. Pada proses normalisasi menggunakan teknik *Gaussian Normalization* atau *Z-Score*, teknik ini mengubah nilai-nilai suatu variabel sehingga distribusi nilai tersebut memiliki rata-rata (mean) 0 dan deviasi standar (standard deviation) 1. Adapun hasil dari proses normalisasi dapat dilihat dari *sample* data pada tabel 5.

Tabel 5. Hasil Normalisasi Variabel

Row ID	Tahun	Hasil Produksi Panen	Luas Panen	Curah Hujan	Kelembapan	Suhu Rata-rata
Row 1	1.673	-0.311	0.231	0.738	0.005	-0.865
Row 2	1.554	0.336	0.208	0.851	0.031	0.389
Row 3	1.434	-0.267	0.166	0.899	0.212	-0.559
Row 4	1.314	-0.236	0.128	0.813	0.297	-0.836
Row 224	1.438	0.385	0.359	0.653	1.202	0.840

Sumber: (Fitri & Nugraha, 2024)

Tabel 5 menunjukkan hasil dataset hasil panen yang telah dilakukan normalisasi, pada Tabel 5 hanya disajikan beberapa data sebagai *sample* dari proses penyelesaian normalisasi. Selanjutnya, proses partitioning yaitu melibatkan pembagian dataset menjadi bagian-bagian yang relevan, yaitu data pelatihan sebesar 70% dan data pengujian 30% dengan metode linear sampling.

### Pengujian Model Prediksi dan Evaluasi

Rangkaian pemodelan pada prediksi dataset hasil panen yang dilakukan pada penelitian ini meliputi pengujian model prediksi *Linear regression*, *Random Forest Regression* serta implementasi ANN dengan arsitektur model *MultiLayer Perceptron*. Pada pengujian model *Linear regression*, *Random Forest Regression* serta *MultiLayer Perceptron*, dataset telah diuji menggunakan model dan proses yang identik. Hasil dari pemrosesan dataset melibatkan pembagian data dengan rasio 70:30 untuk data pelatihan dan pengujian masing-masing sebesar 130 dan 57 instance. Selanjutnya, dilakukan validasi *Cross-Validation* pada data training untuk model pengujian. Pada data pelatihan, model *learner* yang telah diuji pada data training selanjutnya diimplementasikan pada data testing, adapun nilai K pada proses validasi menggunakan K-10 *Cross-*

Validation dengan target variabel yang digunakan yaitu hasil produksi panen, serta teknik yang digunakan yaitu teknik *Stratified Sampling*.

Adapun hasil dari pengujian tiga model prediction, didapatkan hasil matrix evaluasi seperti yang ditampilkan pada tabel 6.

Tabel. 6 Hasil Prediksi Menggunakan Model Prediksi LR, RFR, ANN dengan MLP

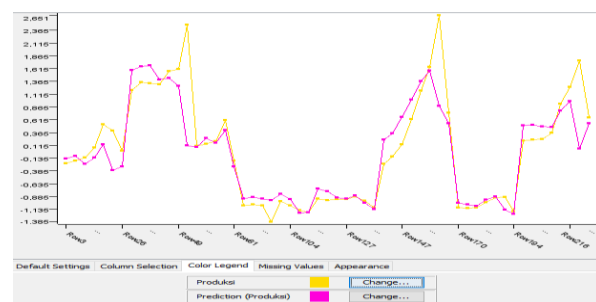
Matrix Evaluasi	LR		RFR		ANN with MLP	
	Training	Testing	Training	Testing	Training	Testing
R <sup>2</sup>	0.85	0.75	0.88	0.90	0.89	0.86
Mean Absolute Error	2	8	3	3	2	7
Mean Squared Error	0.21	0.27	0.22	0.18	0.04	0.04
root Mean Squared Error	4	8	9	6	4	5
mean signed difference	0.14	0.26	0.12	0.07	0.00	0.00
Mean Absolute Percentage Error	1	0	5	9	6	9
adjusted R <sup>2</sup>	0.37	0.51	0.35	0.28	0.08	0.09
Mean Absolute Percentage Error	6	6	3	1	0	4
Mean Absolute Percentage Error	-	-	-	0.02	0.00	-
Mean Absolute Percentage Error	0.00	0.06	0.01	4	1	0.01
Mean Absolute Percentage Error	2	1	2	4	1	5
Mean Absolute Percentage Error	0.72	1.09	0.54	1.05	0.38	0.18
Mean Absolute Percentage Error	4	1	1	3	5	8
adjusted R <sup>2</sup>	0.85	0.75	0.88	0.90	0.89	0.86
R <sup>2</sup>	2	8	3	3	2	7

Sumber: (Fitri & Nugraha, 2024)

Pada Tabel 6 hasil dari pengujian tiga model prediksi, ada data training dan testing memberikan gambaran tentang kinerja masing-masing model. Pada interpretasi dari nilai RMSE ketiga model, Model *Linear regression* memiliki kinerja yang baik pada data training RMSE (0.376), namun tampak ada peningkatan RMSE pada data testing (0.51), yang mungkin menunjukkan adanya overfitting atau ketidakmampuan model untuk generalisasi pada data yang belum pernah dilihat. Lalu Model *Random Forest Regression* menunjukkan kinerja yang baik pada kedua set data. RMSE yang rendah pada data testing (0.281) dapat menunjukkan kemampuan model untuk memberikan prediksi yang lebih akurat pada data yang belum pernah dilihat sebelumnya. Sedangkan Model MLP menunjukkan kinerja yang sangat baik dengan RMSE yang sangat rendah pada kedua set data training dan testing. Hal ini dapat menunjukkan kemampuan model neural network ini untuk menangkap pola yang kompleks dan memprediksi dengan tingkat akurasi yang tinggi.

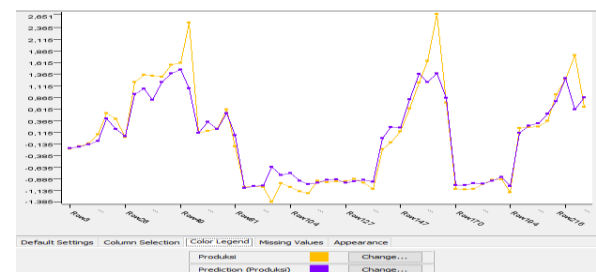
Selanjutnya analisis hasil dari metrik MAE dan MAPE untuk tiga metode yang berbeda pada data training dan testing berdasarkan Tabel 6 ialah Model *Linear regression* memiliki nilai MAE yang

relatif rendah, tetapi terdapat peningkatan antara nilai MAE pada data training dan testing, menunjukkan mungkin adanya overfitting atau ketidakmampuan model untuk menggeneralisasi pada data yang belum pernah dilihat. Lalu Model *Random Forest Regression* menunjukkan hasil yang baik dengan nilai MAE yang rendah pada kedua set data training dan testing. Namun, MAPE pada data testing tetap tinggi, menunjukkan potensi adanya deviasi besar antara prediksi dan nilai aktual. Sedangkan model MLP menunjukkan kinerja yang sangat baik dengan nilai MAE yang sangat rendah pada kedua set data training dan testing. MAPE juga rendah, menunjukkan tingkat kesalahan prediksi yang rendah.



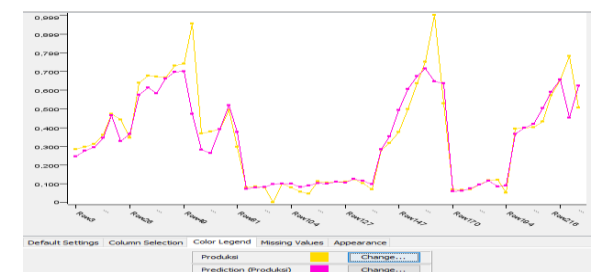
Sumber: (Fitri & Nugraha, 2024)

Gambar 3. Hasil *Line Plot* Prediksi Model *Linear regression*



Sumber: (Fitri & Nugraha, 2024)

Gambar 4. Hasil *Line Plot* Prediksi Model *Random Forest Regression*



Sumber: (Fitri & Nugraha, 2024)

Gambar 5. Hasil *Line Plot* Prediksi Model *MultiLayer Perceptron*

Pada Gambar 4, Gambar 5 serta Gambar 6 merupakan visualisasi hasil prediksi dari tiga model prediction, dimana dari masing-masing gambar menunjukkan ketepatan dalam

memprediksi hasil panen sesuai dengan hasil analisa Tabel 6.

### KESIMPULAN

Berdasarkan hasil prediksi pada data *testing*, model terbaik untuk meramalkan hasil panen adalah *MultiLayer Perceptron* (MLP). Model MLP menunjukkan kinerja terbaik dengan tingkat kesalahan sekitar 0.080. Sebagai perbandingan, model *Linear regression* dan *Random Forest Regression* memiliki tingkat kesalahan yang lebih tinggi, masing-masing sekitar 0.510 dan 0.281. Dengan demikian, prediksi hasil panen yang paling akurat diperoleh melalui penggunaan model *MultiLayer Perceptron* (MLP) dengan tingkat kesalahan sekitar 0.080. Seiring dengan penerapan model *MultiLayer Perceptron* (MLP) yang telah menunjukkan kinerja terbaik dalam meramalkan hasil panen. Pada penelitian selanjutnya dapat dilakukan penyesuaian *hyperparameter* seperti *learning rate* dan jumlah *neuron*, pemantauan kinerja secara berkala dengan dataset validasi, evaluasi fitur yang relevan, dan pertimbangan penggunaan data tambahan untuk mendapatkan hasil akurasi yang lebih baik.

### REFERENSI

- Alita, D., Putra, A. D., & Darwis, D. (2021). Analysis of classic assumption test and multiple *Linear regression* coefficient test for employee structural office recommendation. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3), 295.
- Dardanella, D., Hidayat, A. P., Santosa, S. H., & Siskandar, R. (2022). Peramalan Harga Jual Cabai Merah Di Pasar Rakyat Kemang Perusahaan Umum Daerah Pasar Tohaga Kabupaten Bogor. *Indonesian Journal of Science Learning*, 3(1), 16–23.
- Evita, C. (2021). Penerapan *Artificial Neural Network* Algoritma Backpropagation Pada Prediksi Produksi Jagung. *Seminar Nasional Fortei Regional 7*, 179–184.
- Fitri, E. (2023). Analisis Perbandingan Metode Regresi Linier, *Random Forest Regression* dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah. *Journal of Applied Computer Science and Technology*, 4(1), 58–64.
- Herlina, N., & Prasetyorini, A. (2020). Effect of Climate Change on Planting Season and Productivity of Maize (*Zea mays* L.) in Malang Regency. *Jurnal Ilmu Pertanian Indonesia*, 25(1), 118–128.
- Herwanto, H. W., Widiyaningtyas, T., & Indriana, P. (2019). Penerapan Algoritme *Linear regression* untuk Prediksi Hasil Panen Tanaman Padi. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 8(4), 364.
- Milniadi, A. D., & Adiwijaya, N. O. (2023). Analisis Perbandingan Model Arima Dan Lstm Dalam Peramalan Harga Penutupan Saham (Studi Kasus : 6 Kriteria Kategori Saham Menurut Peter Lynch). *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, Dan Pendidikan*, 2(6), 1683–1692.
- Nuraisah, G., & Kusumo, R. A. B. (2019). Dampak perubahan iklim terhadap usahatani padi di desa Wanguk kecamatan Anjatan kabupaten Indramayu. *Mimbar Agribisnis: Jurnal Pemikiran Masyarakat Ilmiah Berwawasan Agribisnis*, 5(1), 60–71.
- Putra, H., & Walmi, N. U. (2020). Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 6(2), 100–107.
- Rasna, & Matdoan, M. R. I. (2022). Metode Bayesian dan Multilayer Perceptron dalam Mengklasifikasi Diabetes Mellitus. *Jurnal Sistem Informasi Dan Teknologi*, 4, 82–86.
- Satria, A., Maulida Badri, R., & Safitri, I. (2023). Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode *Machine learning*. *Digital Transformation Technology (Digitech) / E*, 3(2), 389–398.
- Sumadi, M. I. T. B. N., Putra, R., & Firmansyah, A. (2022). Peran Perkembangan Teknologi Pada Profesi Akuntan Dalam Menghadapi Industri 4.0 Dan Society 5.0. *Journal of Law, Administration, and Social Science*, 2(1), 56–68.
- Wardhani, R., Nafiiyah, N., & Haydar, M. A. (2022). Algoritma Deep Learning dalam Memprediksi Hasil Panen Padi di Kabupaten Lamongan. *Jurnal Informatika: Jurnal Pengembangan IT*, 7(1), 13–17.
- Wibawa, A. P., Lestari, W., Utama, A. B. P., Saputra, I. T., & Izdihar, Z. N. (2020). Multilayer Perceptron untuk Prediksi Sessions pada Sebuah Website Journal Elektronik. *Indonesian Journal of Data and Science*, 1(3), 57–67.
- Zulvian, S. A., Prihandani, K., & Ridha, A. A. (2021). Perbandingan Metode Msd Dan Cosine Similarity Pada Sistem Rekomendasi Item-Based Collaborative Filtering Comparison of Msd and Cosine Similarity Methods in the Item-Based Collaborative Filtering Recommendation System. *Journal of Information Technology and Computer Science (INTECOMS)*, 4(2), 2021.