

# Penerapan Artificial Neural Network untuk Prediksi Produksi Padi di Sumatera

Setiawan Cahyono Putro<sup>1</sup>, Muhammad Imron Rosadi<sup>2</sup>

<sup>1,2</sup>Jurusan Teknik Informatika, Fakultas Teknik, Universitas Yudharta, Indonesia

<sup>1</sup>setiawancahyo01@gmail.com, <sup>2</sup>imron.rosadi@yudharta.ac.id

---

## Abstrak

Produksi padi memainkan peran penting dalam menjaga ketahanan pangan nasional, khususnya di wilayah Sumatera yang merupakan salah satu lumbung padi utama di Indonesia. Namun, prediksi produksi padi di wilayah ini masih menghadapi tantangan akibat keterbatasan metode konvensional seperti regresi linear yang kurang mampu menangkap hubungan non-linier antara variabel agriklim dan hasil produksi. Masalah spesifik yang diangkat dalam penelitian ini adalah belum tersedianya model prediksi yang mampu menyesuaikan pola agriklim dan produksi secara lokal di tiap provinsi, mengingat karakteristik iklim dan pertanian yang sangat bervariasi antar wilayah. Penelitian ini bertujuan untuk membangun model prediksi produksi padi menggunakan metode *Artificial Neural Network* (ANN) yang dilatih secara terpisah untuk setiap provinsi di Sumatera. Dataset yang digunakan bersifat deret waktu (*time-series*) dari tahun 1993 hingga 2020, mencakup delapan provinsi dengan fitur utama meliputi curah hujan, suhu rata-rata, kelembapan, dan luas panen. Pra-pemrosesan dilakukan melalui penanganan missing value, outlier, normalisasi data, dan seleksi fitur dengan analisis multikolinearitas (VIF). Model ANN dibangun menggunakan algoritma *MLPRegressor* dengan konfigurasi multi-layer dan fungsi aktivasi tanh. Evaluasi performa dilakukan menggunakan *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE). Hasil penelitian menunjukkan bahwa ANN memberikan nilai MSE 0.0138 dan nilai MAE 0.0489 yang lebih rendah dibandingkan metode regresi linear dan GLM, menunjukkan bahwa ANN lebih efektif dalam menangkap pola data yang kompleks dan variatif untuk prediksi produksi padi di Sumatera.

**Kata kunci:** *Artificial Neural Network*, Prediksi Padi, MSE, *Multicollinearity*, Sumatera

---

## 1. Pendahuluan

Produksi padi memainkan peran strategis dalam menjaga ketahanan pangan di Indonesia. Peningkatan produktivitas padi sangat penting untuk mengimbangi pertumbuhan penduduk dan mengurangi ketergantungan pada impor beras (Marwanti Marwanti et al., 2023), (Azhari, 2026). Pulau Sumatera, sebagai salah satu wilayah penghasil padi terbesar, memiliki kontribusi signifikan dalam memenuhi kebutuhan pangan nasional. Namun, produksi padi di Sumatera sangat dipengaruhi oleh faktor iklim, seperti curah hujan, suhu udara, kelembapan, dan perubahan luas panen yang dinamis setiap tahunnya (Mayasari, Nugraha, Juwita, & Heryana, 2023).

Sebagian besar pendekatan prediksi produksi padi di Indonesia, khususnya di wilayah Sumatera, masih didominasi oleh penggunaan metode regresi linear. Metode ini memiliki keterbatasan dalam menangkap pola non-linier antar variabel agriklim dan produksi padi (Adin Musababa, 2024), (Wulandari & Rumini, 2023). Studi terbaru menunjukkan bahwa pendekatan regresi tidak cukup adaptif terhadap kompleksitas dan variabilitas data produksi padi yang tinggi ketika digunakan untuk wilayah geografis yang luas seperti Sumatera. Hal ini menyebabkan rendahnya akurasi prediksi dan

berpotensi menimbulkan kesalahan dalam pengambilan keputusan strategis oleh pemangku kepentingan (Loban, 2023).

Artificial Neural Network (ANN) merupakan metode pembelajaran mesin yang dirancang untuk mengenali pola kompleks dan hubungan non-linier dalam data. ANN telah banyak digunakan untuk berbagai aplikasi prediksi di sektor pertanian dan terbukti menghasilkan akurasi yang tinggi (Putra & Ulfa, 2020), (Tbn & R., 2024). Keunggulan ANN dalam mengolah data dengan variabel input yang beragam dan tidak saling linier menjadikannya metode yang tepat untuk diterapkan pada prediksi produksi padi berbasis parameter iklim dan pertanian (Amri, Ramadhan, Ainurrofiah, & Haris, 2023).

Meskipun telah banyak studi mengenai penerapan ANN di bidang pertanian, kebanyakan masih dilakukan pada skala terbatas, seperti satu provinsi atau kabupaten (Wardhani, Nafiyah, & Haydar, 2022). Hingga saat ini, belum banyak ditemukan penelitian yang menerapkan model ANN untuk prediksi produksi padi secara komprehensif di seluruh wilayah provinsi di Sumatera. Pendekatan lintas-provinsi dapat memberikan gambaran prediktif yang lebih luas dan membantu perencanaan produksi pangan nasional secara makro (Ramadhona, Setiawan, & Bachtiar, 2018), (Satria, Badri, & Safitri, 2023).

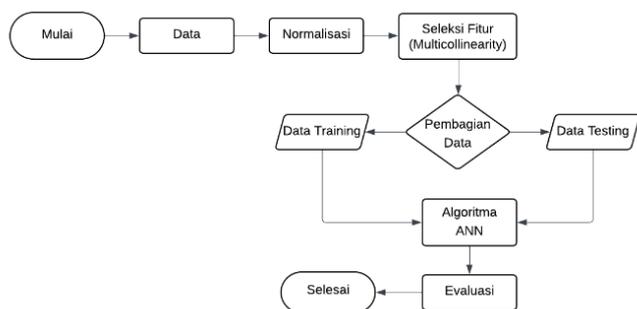
Namun demikian, pemanfaatan Artificial Neural Network (ANN) untuk prediksi produksi padi di wilayah Sumatera secara terpisah per provinsi masih jarang dilakukan, terutama dengan mempertimbangkan karakteristik agroklim yang berbeda-beda di setiap daerah. Studi-studi terdahulu umumnya masih terbatas dalam ruang lingkup geografis dan cenderung menggunakan metode prediktif konvensional seperti regresi linear, yang memiliki keterbatasan dalam mengenali pola hubungan non-linier pada data produksi padi. Hal ini menimbulkan kesenjangan dalam pengembangan model prediksi yang mampu menyesuaikan kompleksitas data di setiap provinsi (Nababan & Nugraha, 2024).

Berdasarkan hal tersebut, pada penelitian ini diarahkan untuk membangun model prediksi produksi padi berbasis Artificial Neural Network (ANN) yang disesuaikan per provinsi di Sumatera dengan mempertimbangkan variabel agroklim seperti curah hujan, suhu udara, kelembapan, dan luas panen. Serta, bagaimana performa model ANN tersebut jika dibandingkan dengan metode Regresi Linear dan Generalized Linear Model (GLM).

Pada penelitian ini setiap model akan dianalisis menggunakan metrik evaluasi seperti Mean Squared Error (MSE) dan Mean Absolute Error (MAE) untuk menilai tingkat akurasi prediksi. Maka dari itu, model ini dapat dijadikan pertimbangan dalam pengambil kebijakan dalam merancang strategi pertanian yang lebih tepat sasaran dan adaptif terhadap kondisi iklim dan sumber daya lahan di Sumatera.

## 2. Metode

Dalam penelitian ini, kami menerapkan serangkaian langkah sistematis untuk menganalisis dan memprediksi produksi padi di Pulau Sumatera menggunakan teknik analisis prediktif. Secara garis besar penelitian ini terdiri dari beberapa tahapan yang dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Pada Gambar 1 ditunjukkan gambar dari alur penelitian yang dimulai dari tahap pengumpulan data, normalisasi data, seleksi fitur, pembagian dataset menjadi data latih dan data uji, penggunaan algoritma dan tahap evaluasi.

## 2.1. Pengumpulan Data

Data yang digunakan pada penelitian ini diperoleh dari platform kaggle <https://www.kaggle.com/datasets/ardikasatria/datase-tanamanpadisumatera> yang merupakan situs publik yang menyediakan berbagai macam dataset. Dataset ini disusun dalam bentuk *time-series*, yang mencakup kurun waktu antara tahun 1993 hingga 2020, dan memuat informasi terkait parameter-parameter seperti provinsi, tahun, luas panen, produksi, curah hujan, suhu rata-rata, dan kelembapan, dari berbagai provinsi di wilayah Pulau Sumatera.

Dataset tersebut mencakup 224 baris data yang terdistribusi dari delapan provinsi di Sumatera, yakni, Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, dan Lampung. Setiap entri dalam dataset merepresentasikan satu titik data produksi padi dalam satu tahun pada suatu provinsi tertentu yang menjadikannya representatif untuk analisis temporal dan spasial. Dataset yang digunakan dalam penelitian ini dapat ditunjukkan pada Tabel 1.

Provinsi	Tahun	Produksi	Luas Panen	Curah Hujan	Kelembaban	Suhu Rata-rata
Riau	2007	490087	147167	3216.3	85.3	27.1
Riau	2008	494260	147796	3026.2	80.2	27.4
Riau	2009	531429	149423	3390	76.2	27.7
Riau	2010	574864	156088	3390	76.2	27.7
Riau	2011	535788	145242	2405	74.5	27
Riau	2012	512152	144015	2636	76.1	27.3
Riau	2013	434144	118518	2628	76	26
Riau	2014	385475	106037	2343	78.1	27.2
Riau	2015	393917	107546	2048.3	80.5	27.2
Riau	2016	325826	72151	2105.6	83.72	27
Riau	2017	337421	70016	2982.9	84.76	26.7
Riau	2018	266375.5	71448.08	1837.1	71.22	25
Riau	2019	230874	63142.04	1608.3	83.77	26.91
Riau	2020	269344.1	64733.13	2584.9	81.8	25.93

Tabel 1. Dataset Produksi Padi di Sumatera

Dari Tabel 1 dapat diketahui bahwa didalam dataset terdapat berbagai kolom seperti provinsi, tahun, produksi, luas panen, curah hujan, kelembapan dan suhu rata-rata dari tiap provinsi dipulau Sumatera yang menjadi variabel untuk eksekusi algoritma ANN.

## 2.2. Pra-Pemrosesan

Pra-Pemrosesan data adalah langkah yang penting untuk menjamin kualitas data yang bagus. Dalam Pra-Pemrosesan data terdapat beberapa aktivitas seperti penghapusan data yang tidak relevan atau tidak sesuai, penanganan nilai kosong (*missing value*), dan normalisasi data.

Penghapusan data yang tidak relevan dilakukan terhadap entri yang tidak memiliki nilai pada salah satu variabel utama seperti produksi, curah hujan, atau suhu rata-rata. Karena proporsi data kosong sangat kecil, penghapusan baris digunakan sebagai pendekatan praktis agar tidak mengganggu distribusi.

Selain itu, data dengan input kategori atau satuan yang tidak sesuai akan dihapus.

Outlier dideteksi menggunakan pendekatan statistik seperti Interquartile Range (IQR). Nilai-nilai yang berada di luar batas normal ( $1.5 \times IQR$  dari kuartil pertama dan ketiga) diperiksa secara manual, dan hanya nilai-nilai yang jelas merupakan kesalahan input (misalnya suhu  $< 10^\circ C$  atau curah hujan negatif) yang akan dihapus.

Karena dataset bersifat deret waktu (*time-series*) dari tahun 1993 hingga 2020, maka seluruh data disusun secara kronologis berdasarkan tahun agar pola temporal dalam data tetap terjaga selama proses pelatihan dan pengujian model.

Normalisasi data dilakukan menggunakan Min-Max Scaling dengan rentang  $[0,1]$ . Metode ini mempertahankan distribusi asli data tanpa mengubah bentuk sebarannya, sehingga hubungan antar variabel tetap terjaga. Pendekatan ini sangat sesuai untuk Artificial Neural Network (ANN), karena fungsi aktivasi seperti *tanh* atau *sigmoid* lebih efektif ketika nilai input berada dalam rentang terbatas, yang berdampak pada konvergensi pelatihan yang lebih cepat dan stabil. Selain itu, dibandingkan metode normalisasi lain seperti *z-score standardization*, Min-Max Scaling lebih tepat digunakan pada dataset yang tidak memiliki distribusi normal dan pada kasus di mana batas bawah dan batas atas nilai fitur perlu dipertahankan secara eksplisit.

Normalisasi data bertujuan untuk memastikan nilai pada setiap fitur memiliki skala yang sama, dengan menggunakan metode *Min-Max Scaling* agar berada dalam rentang  $[0,1]$ . Hal ini bertujuan untuk menghindari dominasi fitur tertentu pada saat proses pelatihan model. Normalisasi dapat dilakukan dengan rumus yang ditunjukkan pada Persaman (1).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Dimana  $X_{norm}$  adalah nilai hasil normalisasi data antara 0 dan 1,  $X$  adalah nilai data asli yang akan dinormalisasi,  $X_{min}$  adalah nilai minimum dari dataset,  $X_{max}$  adalah nilai maksimum dalam dataset.

Tahapan ini sejalan dengan menekankan bahwa pra-pemrosesan data ini untuk meningkatkan performa model *Machine Learning* (Gupta, Sarkar, Dhakre, & Bhattacharya, 2023). Dibawah ini disajikan table hasil normalisasi yang akan ditunjukkan pada Tabel 2.

Tabel 2. Hasil Normalisasi Dataset

	Produksi	Luas Panen	Curah Hujan	Kelembaban	Suhu Rata-rata
0	0.265928	0.3217	0.265025	0.763736	0.505222
1	0.259761	0.328435	0.245023	0.767033	0.617493
2	0.276958	0.341048	0.236532	0.783516	0.532637
3	0.284445	0.352128	0.251816	0.791209	0.507833
4	0.273893	0.338958	0.21068	0.776374	0.537859
5	0.281439	0.373952	0.234456	0.78022	0.60705
6	0.296761	0.366449	0.293518	0.78544	0.515666
7	0.298455	0.337975	0.330635	1	0.640992

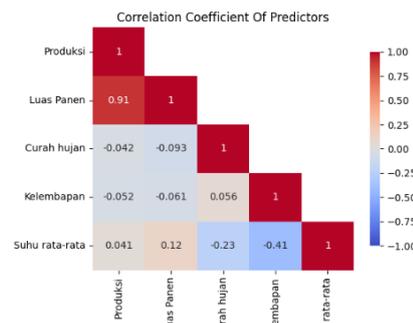
	Produksi	Luas Panen	Curah Hujan	Kelembaban	Suhu Rata-rata
8	0.310979	0.286649	0.276668	0.41978	0.875979
9	0.262751	0.311253	0.202717	0.399725	0.915144
10	0.248788	0.376107	0.242419	0.452198	0.941253
11	0.270312	0.38022	0.165016	0.731868	0.941253
12	0.282900	0.339368	0.092084	0.695055	0.601828
13	0.311925	0.318242	0.05359	0.730769	0.592689
14	0.312913	0.36756	0.224833	0.75	0.546997
15	0.280965	0.328519	0.197736	0.667582	0.627937
16	0.308058	0.365903	0.25559	0.673077	0.614883
17	0.360840	0.35714	0.332767	0.747253	0.640992
18	0.357580	0.392226	0.197283	0.692308	0.640992
19	0.318191	0.401017	0.165204	0.697802	0.614883
20	0.472930	0.439777	0.264383	0.728022	0.627937
21	0.367315	0.386607	0.385301	0.662088	0.640992
22	0.395606	0.491503	0.255213	0.708791	0.640992
23	0.441866	0.284	0.164827	0.8	0.643603
24	0.503495	0.285749	0.317653	0.861813	0.563969
25	0.353246	0.329021	0.227437	0.818132	0.560052
26	0.345483	0.304931	0.322464	0.815934	0.582245
27	0.375893	0.314636	0.263553	0.731319	0.420366

Berdasarkan Tabel 2, terlihat bahwa seluruh nilai variabel telah terstandardisasi sehingga nilai maksimum menjadi 1 dan nilai minimum menjadi 0, dengan nilai lainnya berada di antara kedua batas tersebut. Normalisasi ini bertujuan untuk memastikan setiap variabel memiliki kontribusi yang seimbang dalam model Artificial Neural Network (ANN) serta membantu mempercepat konvergensi pelatihan dan meningkatkan akurasi prediksi.

### 2.3. Seleksi Fitur

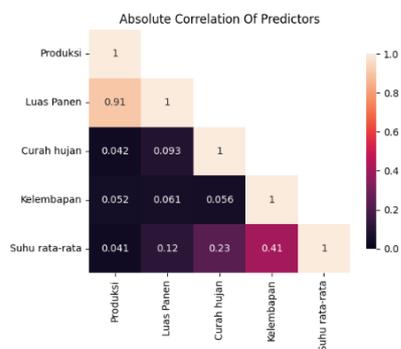
Dalam penelitian ini menggunakan seleksi fitur dengan metode *multicollinearity*, yang bertujuan untuk memastikan bahwa variabel-variabel independen yang digunakan benar-benar relevan dan bebas dari masalah *multicollinearity*. *Multicollinearity* merupakan kondisi ketika dua atau lebih fitur saling berkorelasi tinggi atau rendah, sehingga dapat menyebabkan ketidakstabilan dalam estimasi parameter model, terutama pada pendekatan statistik seperti regresi maupun pada algoritma pembelajaran mesin seperti Artificial Neural Network (ANN).

Untuk mengidentifikasi potensi *multicollinearity*, dilakukan analisis korelasi antar variabel menggunakan matriks korelasi Pearson yang divisualisasikan dalam bentuk heatmap, sebagaimana ditunjukkan pada Gambar 2 dan Gambar 3.



Gambar 2. Hasil Koefisien Multicollinearity

Gambar 2 menunjukkan matriks korelasi antar variabel prediktor dengan variabel target produksi padi. Terlihat bahwa variabel *luas panen* memiliki korelasi paling tinggi terhadap produksi dengan nilai 0,91, yang menandakan hubungan positif kuat antara kedua variabel tersebut. Sementara itu, variabel *curah hujan* dan *kelembapan* menunjukkan korelasi yang sangat rendah terhadap produksi, masing-masing bernilai  $-0,042$  dan  $-0,052$ , sehingga pengaruhnya relatif kecil. Variabel *suhu rata-rata* memiliki korelasi positif rendah terhadap produksi sebesar 0,041, namun berkorelasi negatif dengan kelembapan ( $-0,41$ ). Hasil ini mengindikasikan bahwa tidak semua variabel memiliki hubungan linier yang kuat dengan produksi padi, sehingga diperlukan metode prediksi yang mampu menangkap pola non-linier seperti Artificial Neural Network (ANN).



Gambar 3. Hasil Absolute Multicollinearity

Berdasarkan Gambar 3, terlihat bahwa variabel Luas Panen memiliki korelasi yang sangat tinggi dengan variabel target Produksi ( $r = 0,91$ ) dan juga menunjukkan korelasi kuat terhadap variabel prediktor lainnya. Korelasi tinggi antar variabel prediktor dapat menandakan adanya potensi multikolinearitas yang dapat mempengaruhi stabilitas model. Selain itu, pada *absolute correlation matrix* terlihat bahwa hubungan antar variabel sebagian besar berada pada tingkat rendah hingga sedang, kecuali hubungan antara Luas Panen dan Produksi. Temuan ini menjadi dasar untuk dilakukan pengujian lanjutan menggunakan *Variance Inflation Factor* (VIF) untuk menentukan variabel mana yang perlu dieliminasi pada tahap seleksi fitur.

Perbedaan utama dengan Gambar 2 dan Gambar 3 adalah bahwa Gambar 2 menampilkan nilai korelasi dengan tanda positif/negatif (arah hubungan), sedangkan Gambar 3 menampilkan korelasi absolut (tanpa tanda negatif), sehingga lebih fokus pada kekuatan hubungan antar variabel tanpa memperhatikan arah hubungannya. Hal ini membuat Gambar 3 lebih jelas untuk analisis seleksi fitur karena yang ditekankan adalah besarnya korelasi, bukan arah pengaruh.

Selain itu, pendekatan kuantitatif dilakukan melalui perhitungan *Variance Inflation Factor* (VIF)

untuk menilai derajat multikolinearitas antar fitur. VIF digunakan untuk mengukur seberapa besar varians koefisien estimasi meningkat akibat adanya korelasi antar variabel independen (Kyriazos & Poga, 2023). Penjelasan dari nilai VIF yaitu:

1. Nilai  $VIF = 1$ , artinya tidak ada korelasi
2. Nilai  $1 < VIF \leq 5$ , artinya menunjukkan adanya korelasi tetapi tidak berpengaruh secara signifikan
3. Nilai  $VIF > 5$ , artinya menunjukkan adanya multicollinearity yang dapat berpengaruh terhadap stabilitas pada model estimasi

Maka nilai dari VIF yang optimal berada dalam rentang 1 dan 5. Nilai VIF yang melebihi batas ini dapat mengindikasikan adanya *multicollinearity* yang tinggi, yaitu kondisi di mana variabel-variabel independen saling berkorelasi secara kuat. Multikolinearitas semacam ini berisiko menyebabkan ketidakstabilan dalam estimasi parameter model regresi, serta menyulitkan interpretasi hasil secara statistik.

Untuk mengatasi kondisi ini, beberapa pendekatan yang dapat dilakukan adalah dengan menghapus fitur yang menyebabkan korelasi tinggi, atau menggabungkan fitur-fitur yang berkaitan erat menjadi satu variabel komposit (Kyriazos & Poga, 2023).

Berdasarkan hasil perhitungan VIF pada dataset yang digunakan, seluruh variabel memiliki nilai di bawah 1.3, yang berarti tidak terdapat indikasi multikolinearitas signifikan dalam dataset. Dengan demikian, seluruh fitur dianggap layak untuk digunakan dalam proses pembangunan model prediksi, karena hubungan antar variabel masih berada dalam batas wajar yang dapat diterima secara statistik.

Tabel 3 menyajikan nilai *Variance Inflation Factor* (VIF) sebagai berikut.

Tabel 3. Nilai Variance Inflation Factor (VIF)

	Variabel	VIF
0	Suhu Rata-rata	1.272314
1	Kelembapan	1.201933
2	Curah Hujan	1.062129
3	Luas Panen	1.018731

Berdasarkan Tabel 3, variabel yang memiliki nilai VIF di bawah ambang batas 5 yang dipertahankan untuk membangun model. Hal ini bertujuan agar model Artificial Neural Network (ANN) yang dibangun menjadi lebih stabil, efisien, dan tidak dipengaruhi oleh redundansi informasi antar fitur. Dengan demikian, model dapat lebih fokus pada variabel yang benar-benar relevan dalam memprediksi produksi padi.

#### 2.4. Artificial Neural Network

Kemudian tahap selanjutnya adalah membangun model prediksi menggunakan Artificial

Neural Network (ANN). Sama halnya dengan metode regresi sebelumnya, proses ini diawali dengan pembagian dataset menjadi data training untuk pelatihan model dan data testing untuk mengevaluasi performa pada data yang belum pernah dilihat sebelumnya. Model ANN dalam penelitian ini dibangun secara terpisah untuk masing-masing provinsi yang ada dalam dataset. Secara matematis ANN dapat digambarkan melalui persamaan berikut.

$$y = f(W^{(4)} \cdot f(W^{(3)} \cdot f(W^{(2)} \cdot f(W^{(1)} \cdot x + b^{(1)}) + b^{(2)}) + b^{(3)}) + b^{(4)} \quad (2)$$

di mana  $x$  adalah vektor input,  $W^{(i)}$  dan  $b^{(i)}$  masing-masing adalah bobot dan bias lapisan ke- $i$ , serta  $f = \tanh$ .

Pemilihan metode ANN didasarkan pada kemampuan model ini dalam menangkap hubungan non-linier antar variabel input, yang tidak dapat dilakukan secara efektif oleh model regresi linear. ANN bekerja dengan membentuk jaringan saraf buatan yang terdiri dari beberapa lapisan neuron lapisan input, lapisan tersembunyi (*hidden layer*), hingga lapisan output dengan bobot (*weights*) dan bias, serta menggunakan fungsi aktivasi non-linear seperti  $\tanh$  untuk mengenali pola kompleks dalam data agriklim (Sivaranjani & Vimal, 2023).

Selain fleksibilitas struktural, ANN juga memiliki kemampuan adaptif melalui proses pelatihan yang didukung oleh algoritma optimasi seperti Adam. Model dapat menyesuaikan bobot dan bias secara dinamis berdasarkan pola data yang tersedia, menjadikannya efektif untuk digunakan pada dataset multivariabel berskala provinsi. Dalam konteks prediksi hasil pertanian, beberapa penelitian menunjukkan bahwa ANN mampu memberikan hasil yang lebih stabil dan akurat dibandingkan pendekatan linear, terutama saat jumlah data tidak besar tetapi memiliki kompleksitas tinggi (Amarasinghe et al., 2024).

Dengan mempertimbangkan fleksibilitas jaringan, ketangguhan terhadap variabel non-linier, dan kapasitas generalisasi yang tinggi, maka ANN dipilih sebagai pendekatan utama dalam penelitian ini untuk membangun model prediksi produksi padi per provinsi di wilayah Sumatera.

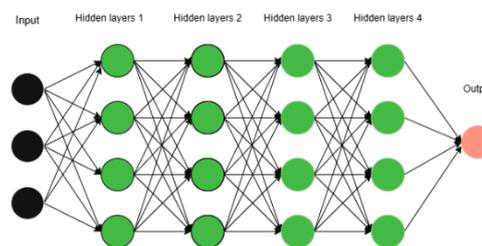
### 3. Hasil dan Pembahasan

#### 3.1 Model Artificial Neural Network

Dalam penelitian ini, model Artificial Neural Network (ANN) dibangun menggunakan algoritma *MLPRegressor*, yang dikonfigurasi dengan arsitektur jaringan berlapis-lapis secara bertingkat, yaitu  $256 \rightarrow 128 \rightarrow 64 \rightarrow 32$  neuron pada *hidden layer*. Untuk mendukung kemampuan model dalam mengenali pola non-linear yang kompleks, digunakan fungsi aktivasi  $\tanh$  pada setiap neuron. Proses pelatihan model dilakukan dengan *optimizer Adam* yang

menggunakan *adaptive learning rate* sebesar 0.0005, sehingga dapat menyesuaikan laju pembelajaran secara otomatis selama proses pelatihan berlangsung. Model dijalankan hingga maksimal 5000 *epoch*, dengan toleransi error (*tol*) sebesar  $1e-6$ , dan tanpa menggunakan teknik *early-stopping*, yang artinya pelatihan akan terus berlanjut hingga salah satu batas tercapai. Proses pelatihan ANN melibatkan dua tahap utama, yaitu data input dialirkan melalui setiap lapisan jaringan untuk menghasilkan output prediksi, dan kemudian menghitung selisih antara nilai prediksi dan target menggunakan *Mean Squared Error (MSE)* dan *Mean Absolute Error (MAE)* untuk memperbarui bobot jaringan secara otomatis agar kesalahan diminimalkan.

Pada Gambar 4 disajikan ilustrasi dari algoritma *Artificial Neural Network*.



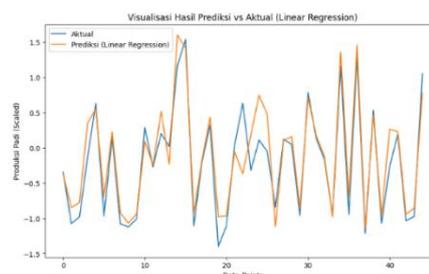
Gambar 4. Ilustrasi Cara Kerja Neural Network

Pada Gambar 4, diilustrasikan bahwa ANN terdiri dari tiga lapisan utama, yaitu *input layer*, *hidden layers*, dan *output layer*. Lapisan input menerima variabel agriklim dan pertanian seperti curah hujan, kelembapan, suhu rata-rata, dan luas panen, yang kemudian diproses melalui dua lapisan tersembunyi (*hidden layers*) dengan neuron yang saling terhubung penuh (*fully connected*). Pada setiap neuron, data dikalikan dengan bobot (*weight*), ditambahkan *bias*, lalu diproses melalui fungsi aktivasi non-linear untuk mengenali pola kompleks dalam data. Lapisan output menghasilkan satu nilai prediksi produksi padi.

#### 3.2. Hasil

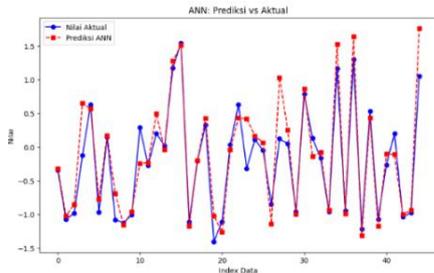
Penelitian ini menggunakan model ANN untuk semua provinsi yang ada pada dataset. Setelah dibangun model prediksi menggunakan ANN kemudian dilakukan uji model ANN.

Dibawah ini disajikan visualisasi antara aktual dan prediksi dari 3 metode yang digunakan, visualisasi ditunjukkan sebagai berikut.



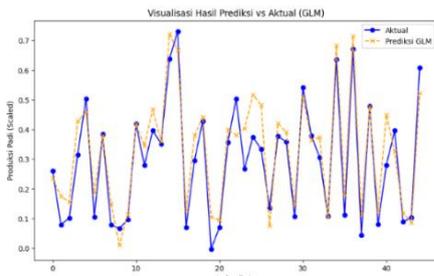
Gambar 5. Visualisasi aktual vs Prediksi *Linear Regression*

Pada Gambar 5, terlihat bahwa hasil prediksi dari *Linear Regression* cenderung kurang mampu mengikuti fluktuasi nilai aktual, terutama pada data dengan perubahan yang tajam. Hal ini menunjukkan keterbatasan model regresi dalam menangkap hubungan non-linear.



Gambar 6. Visualisasi aktual vs Prediksi ANN

Sebaliknya, Gambar 6 menunjukkan bahwa prediksi menggunakan ANN jauh lebih mendekati nilai aktual. Kurva prediksi ANN mengikuti pola naik-turun dari data aktual secara lebih konsisten, mencerminkan kemampuan model dalam mengenali struktur kompleks dan hubungan non-linier dalam data agriklim.



Gambar 7. Visualisasi aktual vs Prediksi GLM

Sementara itu, Gambar 7 menampilkan hasil prediksi dari GLM yang secara visual juga menunjukkan adanya deviasi signifikan terhadap nilai aktual pada beberapa titik, meskipun dalam beberapa segmen model cukup akurat. Secara umum, visualisasi ini mengonfirmasi hasil evaluasi kuantitatif bahwa ANN memberikan performa prediksi terbaik di antara ketiga metode yang diuji.

Kemudian untuk mengukur kemampuan prediksi akan ditampilkan hasil uji model ANN ini dilakukan dengan menggunakan metode *Mean Squared Error (MSE)* dan *Mean Absolute Error (MAE)* seperti pada persamaan berikut.

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

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x| \tag{4}$$

Di mana  $n$  adalah jumlah data,  $Y_i$  adalah nilai aktual,  $Y' i$  adalah nilai prediksi. Teknik ini akan mengukur rata-rata dari kuadrat selisih antara nilai yang diprediksi dan nilai groundtruth.

Nilai Mean Squared Error (MSE) dan Mean Absolute Error (MAE) digunakan sebagai metrik evaluasi performa model prediksi untuk mengukur seberapa besar perbedaan antara nilai prediksi dan nilai aktual pada dataset. MSE mengukur rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi. Metrik ini sensitif terhadap outlier karena kesalahan dipangkatkan dua, sehingga kesalahan besar memiliki dampak yang lebih signifikan. nilai MSE yang lebih kecil menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang rendah. Kemudian MAE mengukur rata-rata selisih absolut antara nilai aktual dan prediksi. Berbeda dengan MSE, MAE tidak terlalu dipengaruhi oleh outlier dan memberikan interpretasi yang lebih mudah karena nilainya berada pada skala yang sama dengan data aslinya.

Rentang performa yang dianggap baik untuk MSE dan MAE bergantung pada skala data. Secara umum, semakin mendekati 0, semakin baik kinerja model, karena hal itu menunjukkan bahwa prediksi hampir sama dengan nilai aktual. Dalam konteks penelitian ini, nilai MSE di bawah 0,02 dan MAE di bawah 0,05 dapat dikategorikan sangat baik, mengingat data telah dinormalisasi dengan skala [0,1].

Hasil Evaluasi model tiap provinsi dapat dilihat pada Tabel 4.

Tabel 4. Perbandingan Hasil Evaluasi Setiap Model Dengan MSE

Provinsi	Mean Squared Error (MSE)		
	ANN	Linear Regression	Generalized Linear Model
Aceh	0.005	0.05	0.06
Sumatera Utara	0.004	0.03	0.04
Sumatera Barat	0.008	0.04	0.06
Riau	0.010	0.01	0.03
Jambi	0.03	0.03	0.04
Sumatera Selatan	0.03	0.03	0.07
Bengkulu	0.011	0.04	0.06
Lampung	0.013	0.05	0.07
<b>Rata - Rata</b>	<b>0.0138</b>	<b>0.035</b>	<b>0.053</b>

Berdasarkan Tabel 4, hasil evaluasi model ANN per provinsi, ditemukan bahwa nilai *Mean Squared Error (MSE)* di Provinsi Aceh dan Sumatera Utara cenderung lebih rendah dibandingkan dengan provinsi lain seperti Jambi dan Sumatera Selatan, yang menunjukkan nilai MSE relatif lebih tinggi. Salah satu penyebab yang memungkinkan hal tersebut muncul adalah jumlah dan kualitas data historis yang tidak merata antar provinsi. Berdasarkan pengamatan terhadap dataset, provinsi Aceh dan Sumatera Utara memiliki jumlah data yang relatif lebih lengkap dan stabil setiap tahunnya, dengan variasi agriklim dan produksi yang konsisten. Hal ini membantu model dalam belajar pola prediksi secara lebih akurat.

Sebaliknya, pada provinsi seperti Jambi dan Sumatera Selatan, terdapat lebih banyak fluktuasi nilai serta beberapa indikasi adanya outlier pada fitur-fitur seperti curah hujan dan suhu rata-rata.

Keberadaan nilai ekstrim tersebut berpotensi menyebabkan gangguan dalam proses pelatihan model ANN, karena meskipun ANN dapat menangani non-linearitas, ia tetap sensitif terhadap noise atau nilai ekstrim dalam data jika tidak ditangani dengan baik pada tahap pra-pemrosesan.

Selain itu, perbedaan struktur lahan, sistem pertanian, atau ketidakteraturan pencatatan data produksi per tahun juga dapat memengaruhi heterogenitas pola agriklim antar provinsi. Model ANN yang dilatih secara lokal untuk setiap provinsi cenderung lebih optimal pada dataset yang konsisten dan representatif, sedangkan pada provinsi dengan data yang lebih sporadis, model mengalami kesulitan dalam menemukan representasi pola yang stabil, sehingga nilai MSE menjadi lebih tinggi.

Tabel 5. Perbandingan Hasil Evaluasi Setiap Model Dengan MAE

Provinsi	Mean Absolute Error (MAE)		
	ANN	Linear Regression	Generalized Linear Model
Aceh	0.012	0.18	0.20
Sumatera Utara	0.014	0.13	0.15
Sumatera Barat	0.023	0.16	0.20
Riau	0.023	0.083	0.13
Jambi	0.115	0.17	0.12
Sumatera Selatan	0.152	0.12	0.22
Bengkulu	0.021	0.12	0.21
Lampung	0.027	0.16	0.22
<b>Rata - Rata</b>	<b>0.0489</b>	<b>0.141</b>	<b>0.185</b>

Berdasarkan Tabel 5, Mean Absolute Error (MAE), metode Artificial Neural Network (ANN) menunjukkan performa terbaik dengan rata-rata MAE sebesar 0,0489, jauh lebih rendah dibandingkan metode Linear Regression sebesar 0,141 dan Generalized Linear Model (GLM) sebesar 0,185.

Berdasarkan hasil evaluasi kinerja model prediksi menggunakan metrik *Mean Squared Error (MSE)* dan *Mean Absolute Error (MAE)*, diperoleh perbandingan performa antara tiga metode yaitu Artificial Neural Network (ANN), Regresi Linear, dan Generalized Linear Model (GLM). Hasil perhitungan menunjukkan bahwa metode ANN memberikan nilai rata-rata MSE dan MAE paling rendah, yaitu MSE 0,0138 dan MAE 0.0489, dibandingkan dengan Regresi Linear dan GLM.

Perbedaan hasil evaluasi matriks MSE dan MAE ini dapat dijelaskan melalui karakteristik data dan kemampuan masing-masing model dalam menangkap pola dalam data. Pertama, data yang digunakan dalam penelitian ini terdiri dari variabel-variabel agriklim seperti curah hujan, suhu rata-rata, kelembapan, dan luas panen, yang memiliki hubungan non-linier dan dinamis terhadap produksi padi. Model regresi linear dan GLM, meskipun efisien dan mudah diinterpretasikan, dibatasi oleh asumsi linearitas, sehingga kurang mampu menangkap interaksi kompleks antar variabel.

Sebaliknya, ANN tidak memerlukan asumsi linearitas, dan dapat memodelkan hubungan non-linier yang lebih dalam melalui jaringan lapisan tersembunyi (hidden layers). ANN mampu menyesuaikan bobot antar neuron berdasarkan pola tersembunyi dalam data, sehingga dapat menggeneralisasi dengan lebih baik terhadap variasi input yang kompleks, terutama pada data per provinsi yang memiliki karakteristik agriklim berbeda-beda.

Temuan ini menunjukkan bahwa ANN lebih sesuai digunakan pada dataset dengan karakteristik kompleks dan hubungan non-linier antar variabel input, seperti halnya data agriklim yang digunakan dalam penelitian ini. Sebaliknya, Regresi Linear dan GLM lebih cocok diterapkan pada data yang lebih sederhana atau ketika jumlah sampel terbatas, meskipun akurasi prediksinya tidak sebaik ANN. Namun perbedaannya tidak terlalu signifikan pada beberapa wilayah, sehingga kedua pendekatan tersebut masih relevan digunakan dalam kondisi tertentu.

#### 4. Kesimpulan

Penelitian ini berhasil membangun model prediksi produksi padi di Sumatera menggunakan Artificial Neural Network (ANN) berbasis data agriklim. Hasil evaluasi model menunjukkan bahwa ANN memberikan performa terbaik dengan nilai rata-rata MSE terendah dibandingkan metode regresi linear dan GLM. Hal ini membuktikan bahwa ANN memiliki keunggulan dalam memodelkan hubungan non-linier antar variabel agriklim seperti curah hujan, suhu, kelembapan, dan luas panen. Model ANN yang dibangun mampu menyesuaikan karakteristik masing-masing provinsi, sehingga memberikan hasil prediksi yang lebih akurat.

Ke depan, model ini dapat dikembangkan lebih lanjut dengan menambahkan variabel lain seperti jenis varietas padi, pemupukan, dan data satelit untuk meningkatkan akurasi. Selain itu, implementasi real-time dan integrasi ke dalam sistem pengambilan keputusan dapat memperkuat kontribusi model terhadap kebijakan ketahanan pangan nasional.

#### Daftar Pustaka:

Adin Musababa, M. (2024). Implementasi Algoritma Linear Regression untuk Prediksi Produksi Tanaman Padi di Kabupaten Grobogan. *Data Sciences Indonesia (DSI)*, 3(2), 68–78. <https://doi.org/10.47709/dsi.v3i2.3118>

Amarasinghe, A., Sangarasekara, I., Silva, N. De, Ariyaratne, M., Amarasinghe, R., Bogahawatte, J., ... Herath, D. (2024). Advancing food sustainability: a case study on improving rice yield prediction in Sri Lanka using weather-based, feature-engineered machine learning models. *Discover Applied Sciences*, 6(11). <https://doi.org/10.1007/s42452-024-06300-7>

- Amri, I. F., Ramadhan, W. N., Ainurrofiah, S., & Haris, M. Al. (2023). Pemodelan ARIMA dan ARIMAX untuk Memprediksi Jumlah Produksi Padi di Kota Magelang. *Square : Journal of Mathematics and Mathematics Education*, 5(2), 93–105. <https://doi.org/10.21580/square.2023.5.2.17059>
- Azhari, D. (2026). Produktivitas dan Prediksi Produksi Padi di Provinsi Bengkulu Productivity and Production Prediction of Paddy in Bengkulu Province, 5(2), 279–288.
- Gupta, A., Sarkar, K., Dhakre, D., & Bhattacharya, D. (2023). Weather based crop yield prediction using artificial neural networks: A comparative study with other approaches. *Mausam*, 74(3), 825–832. <https://doi.org/10.54302/mausam.v74i3.174>
- Kyriazos, T., & Poga, M. (2023). Dealing with Multicollinearity in Factor Analysis: The Problem, Detections, and Solutions. *Open Journal of Statistics*, 13(03), 404–424. <https://doi.org/10.4236/ojs.2023.133020>
- Loban, J. M. (2023). Analisis Regresi Faktor-Faktor Yang Mempengaruhi Hasil Produksi Padi Di Indonesia Bagian Barat. *EduMatSains : Jurnal Pendidikan, Matematika dan Sains*, 8(1), 138–145. <https://doi.org/10.33541/edumatsains.v8i1.4856>
- Marwanti Marwanti, Setyono Hari Adi, Hendri Sosiawan, Muhrizal Sarwani, Gatot Irianto, & Mohammad Ismail Wahab. (2023). Disrupsi Sistem Produksi Padi Nasional: Mampukah Indonesia Memenuhi Kebutuhan Beras di Tahun 2045? *Jurnal Triton*, 14(2), 403–421. <https://doi.org/10.47687/jt.v14i2.588>
- Mayasari, R., Nugraha, B., Juwita, A. R., & Heryana, N. (2023). Analisis Produktifitas Padi di Pulau Sumatera menggunakan Exploratory Data Analysis ( EDA ). *Jurnal Elektronik Sistem Informasi Unsika*, 1(1), 17–24.
- Nababan, Y., & Nugraha, I. (2024). Penerapan Data Mining Produksi Padi di Pulau Sumatera Menggunakan Analisis Regresi Linear. *Jurnal Teknik Industri Terintegrasi*, 7(1), 262–272. <https://doi.org/10.31004/jutin.v7i1.23545>
- Putra, H., & Ulfa, N. (2020). Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation. *Nasional Teknologi dan Sistem Informasi*, 02, 100–107.
- Ramadhona, G., Setiawan, B. D., & Bachtiar, F. A. (2018). Prediksi Produktivitas Padi Menggunakan Jaringan Syaraf Tiruan Backpropagation. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(12), 6048–6057.
- Satria, A., Badri, R. M., & Safitri, I. (2023). Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning. *Digital Transformation Technology*, 3(2), 389–398. <https://doi.org/10.47709/digitech.v3i2.2852>
- Sivaranjani, T., & Vimal, S. P. (2023). AI Method for Improving Crop Yield Prediction Accuracy Using ANN. *Computer Systems Science and Engineering*, 47(1), 154–170. <https://doi.org/10.32604/csse.2023.036724>
- Tbn, A. F. A., & R., R. K. (2024). Penerapan Algoritma Backpropagation untuk Prediksi Hasil Panen Padi di Kabupaten Labuhan Batu Utara. *Jurnal Teknologi Sistem Informasi dan Aplikasi*, 7(1), 335–342. <https://doi.org/10.32493/jtsi.v7i1.38318>
- Wardhani, R., Nafiyah, 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. <https://doi.org/10.30591/jpit.v7i1.2581>
- Wulandari, D., & Rumini, R. (2023). Pemodelan dan Prediksi Produksi Padi Menggunakan Regresi Linear. *Smart Comp: Jurnalnya Orang Pintar Komputer*, 12(4). <https://doi.org/10.30591/smartcomp.v12i4.5905>
- Nababan, Y., & Nugraha, I. (2024). Penerapan Data Mining Produksi Padi di Pulau Sumatera Menggunakan Analisis Regresi Linear. *Jurnal Teknik Industri Terintegrasi*, 7(1), 262–272. <https://doi.org/10.31004/jutin.v7i1.23545>