

**PREDIKSI HASIL PANEN PADI MENGGUNAKAN *ARTIFICIAL  
NEURAL NETWORK* BERDASARKAN FAKTOR CUACA  
DAN KONDISI TANAH**

**TESIS**

**Oleh:  
PRIYANTO  
NIM. 240605210013**



**PROGRAM MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

**PREDIKSI HASIL PANEN PADI MENGGUNAKAN *ARTIFICIAL  
NEURAL NETWORK* BERDASARKAN FAKTOR CUACA  
DAN KONDISI TANAH**

**TESIS**

**Diajukan kepada:**

**Universitas Islam Negeri Maulana Malik Ibrahim Malang  
untuk Memenuhi Salah Satu Persyaratan dalam Memperoleh Gelar  
Magister Komputer (M. Kom)**

**Oleh:**

**PRIYANTO  
NIM. 240605210013**

**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

**PREDIKSI HASIL PANEN PADI MENGGUNAKAN *ARTIFICIAL  
NEURAL NETWORK* BERDASARKAN FAKTOR CUACA  
DAN KONDISI TANAH**

**TESIS**

**Diajukan kepada:**

**Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
untuk Memenuhi Salah Satu Persyaratan dalam Memperoleh Gelar  
Magister Komputer (M. Kom)**

**Oleh:**

**PRIYANTO  
NIM. 240605210013**

**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Priyanto  
NIM : 240605210013  
Program Studi : Magister Informatika  
Fakultas : Sains dan Teknologi  
Judul Thesis : “Prediksi Hasil Panen Padi Menggunakan *Artificial Neural Network* Berdasarkan Faktor Cuaca Dan Kondisi Tanah”

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-banar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil plagiasi, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 02 Desember 2025

Yang Membuat Pernyataan



Priyanto

NIM. 240605210013

**HALAMAN PERSETUJUAN UJIAN THESIS**

**PREDIKSI HASIL PANEN PADI MENGGUNAKAN *ARTIFICIAL NEURAL NETWORK* BERDASARKAN FAKTOR CUACA DAN KONDISI TANAH**

**THESIS**

**Oleh:  
PRIYANTO  
NIM. 240605210013**

Telah diperiksa dan disetujui untuk diuji :  
Tanggal 18 November 2025

Pembimbing I

Prof. Dr. Ir. Muhammad Faisal, M.T  
NIP.19740510 200501 1 007

Pembimbing II

  
Dr. Mochamad. Imamudin Lc, MA  
NIP. 19740602 200901 1 010

Mengetahui,  
Ketua Program Studi Magister Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Maulana Malik Ibrahim Malang



Prof. Dr. Ir. Muhammad Faisal, M.T  
NIP.19740510 200501 1 007

**HALAMAN PENGESAHAN**  
**PREDIKSI HASIL PANEN PADI MENGGUNAKAN *ARTIFICIAL NEURAL NETWORK* BERDASARKAN FAKTOR CUACA DAN KONDISI TANAH**

**THESIS**

**Oleh:**  
**PRIYANTO**  
**NIM. 240605210013**

Telah Dipertahankan di Depan Dewan Penguji Thesis  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Magister Komputer (M.Kom)  
Tanggal 18 November 2025

**Susunan Dewan Penguji**

Penguji 1	: <u>Dr. Usman Pagalay, M.Si</u> NIP. 19650414 200312 1 001
Penguji 2	: <u>Dr. Ririen Kusumawati, S.Si., M.Kom</u> NIP. 19720309 200501 2 002
Pembimbing 1	: <u>Prof. Dr. Ir. Muhammad Faisal, M.T</u> NIP.19740510 200501 1 007
Pembimbing 2	: <u>Dr. Mochamad. Inamudin Lc, MA</u> NIP. 19740602 200901 1 010

**Tanda Tangan**

(  )  
(  )  
(  )  
(  )

Mengetahui,  
Program Studi Magister Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Maulana Malik Ibrahim Malang

  
Dr. Muhammad Faisal, M.T  
NIP.19740510 200501 1 007

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb.*

Syukur alhamdulillah penulis hanturkan kehadiran Allah SWT SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga dapat menyelesaikan studi di Program Studi Magister Informatika di Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan Thesis ini dengan baik. Oleh karena itu penulis haturkan ucapan terima kasih kepada semua pihak yang telah membantu terselesaikannya Thesis ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Bapak Prof. Dr. Ir. Muhammad Faisal, M.T , selaku Ketua Program Studi Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Bapak Prof. Dr. Ir. Muhammad Faisal, M.T dan Bapak Dr. M. Imamudin Lc., MA selaku dosen pembimbing Thesis, yang telah banyak memberikan pengarahan dan pengalaman yang berharga.
3. Segenap civitas akademika Program Studi Magister Informatika, terutama seluruh Bapak/ Ibu dosen, terima kasih atas segenap ilmu dan bimbingan.
4. Istri dan anak tercinta yang senantiasa memberikan do'a dan semangat kepada saya untuk meraih kesuksesan.
5. Semua rekan-rekan magister informatika angkatan 2024 yang membantu dan menemani bersama-sama di studi magister ini.
6. Semua pihak yang ikut membantu dalam menyelesaikan Thesis ini baik berupa materil maupun moril.

Penulis menyadari bahwa dalam penyusunan Thesis ini masih terdapat kekurangan dan berharap semoga bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. Amin Ya Rabbal Alamin.

*Wassalamu'alaikum Wr. Wb.*

Malang, 10 Desember 2025

Penulis,

## ABSTRAK

Priyanto 2025. **Prediksi Hasil Panen Padi Menggunakan *Artificial Neural Network* Berdasarkan Faktor Cuaca dan Kondisi Tanah**. Thesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Ir. Muhammad Faisal, M.T (II) Dr. Mochamad. Imamudin Lc, MA

Indonesia merupakan salah satu negara agraris terbesar di dunia yang sangat bergantung pada produksi padi sebagai sumber pangan utama bagi penduduknya. Namun, perubahan iklim yang tidak menentu dan variasi kondisi tanah di berbagai daerah menjadi tantangan besar dalam menjamin kestabilan hasil panen padi. Ketidakpastian tersebut mempersulit petani dalam merencanakan waktu tanam dan pengelolaan lahan yang optimal. Oleh karena itu, diperlukan pendekatan yang inovatif dan berbasis teknologi untuk memprediksi hasil panen secara lebih akurat. Penelitian ini bertujuan untuk membangun model prediksi hasil panen padi dengan menggunakan *Artificial Neural Network (ANN)*, khususnya arsitektur *Multilayer Perceptron (MLP)*, yang mampu memodelkan hubungan non-linear antara faktor-faktor lingkungan dengan hasil produksi padi. Model *ANN* dirancang dengan memasukkan variabel-variabel cuaca seperti curah hujan, suhu udara, dan kelembapan, serta kondisi tanah seperti pH, kelembapan tanah, dan kandungan nutrisi sebagai input utama. Data diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta Badan Pusat Statistik (BPS) Wilayah Malang dalam rentang waktu 2009 hingga 2024. Proses *data preprocessing* meliputi normalisasi data dengan metode *Z-Score* serta penanganan nilai hilang untuk memastikan kualitas input yang optimal. Selain itu, proses *hyperparameter tuning* diterapkan menggunakan metode *Grid Search* untuk meningkatkan performa dan stabilitas model, serta meminimalisir risiko *overfitting* dan *underfitting*. Evaluasi dilakukan menggunakan metrik *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *R-Squared (R<sup>2</sup>)*. Hasil yang diperoleh diharapkan tidak hanya memberikan prediksi yang akurat, tetapi juga dapat digunakan sebagai dasar dalam pengambilan keputusan dalam perencanaan ketahanan pangan yang adaptif dan berbasis data. Dengan demikian, penelitian ini berkontribusi pada pengembangan pertanian presisi di Indonesia yang lebih modern, efisien, dan berkelanjutan.

**Kata kunci:** *Artificial Neural Network*, Prediksi hasil panen, Hyperparameter tuning.



## ABSTRACT

Priyanto 2025. **Rice Yield Prediction Using *Artificial Neural Network* Based on Weather Factors and Soil Conditions**. Thesis. Master of Informatics Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Supervisor: (I) Prof. Dr. Ir. Muhammad Faisal, M.T (II) Dr. Mochamad. Imamudin Lc, MA

Indonesia is one of the largest agrarian countries in the world that relies heavily on rice production as the main source of food for its population. However, erratic climate change and variations in soil conditions in various regions are a major challenge in ensuring the stability of rice crops. This uncertainty makes it difficult for farmers to plan optimal planting time and land management. Therefore, an innovative and technology-based approach is needed to predict crop yields more accurately. This study aims to build a predictive model of rice yield using *Artificial Neural Network (ANN)*, especially *Multilayer Perceptron (MLP)* architecture, which is able to model the non-linear relationship between environmental factors and rice production. The *ANN* model is designed by incorporating weather variables such as *Rainfall*, air *Temperature*, and *Humidity*, as well as soil conditions such as pH, soil moisture, and nutrient content as the main inputs. Data was obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG) and the Central Statistics Agency (BPS) of the Malang Region in the period 2009 to 2024. The *data preprocessing* process includes data normalization with the *Z-Score* method as well as missing value handling to ensure optimal input quality. In addition, the *hyperparameter tuning* process is applied using the *Grid Search* method to improve model performance and stability, as well as minimize the risk of *overfitting* and *underfitting*. The evaluation was conducted using the *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, and *R-Squared (R<sup>2</sup>)* metrics. The results obtained are expected not only to provide accurate predictions, but also to be used as a basis for decision-making in adaptive and data-driven food security pLANNing. Thus, this research contributes to the development of precision agriculture in Indonesia that is more modern, efficient, and sustainable.

**Kata kunci:** *Artificial Neural Network*, Yield prediction, Hyperparameter tuning.

## الملخص

بريانتو 2025. التنبؤ بمحصول الأرز باستخدام الشبكة العصبية الاصطناعية بناءً على العوامل الجوية وظروف التربة. الرسالة الجامعية. برنامج دراسة ماجستير المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية، ملانج. المشرف: (أ) أستاذ د. محمد فيصل (ب) دكتور. محمد إمام الدين ليسا ناس، ماجستير.

إندونيسيا هي واحدة من أكبر البلدان الزراعية في العالم التي تعتمد بشكل كبير على إنتاج الأرز كمصدر رئيسي للغذاء لسكانها. ومع ذلك، فإن تغير المناخ غير المنتظم والتغيرات في ظروف التربة في مختلف المناطق تشكل تحدياً رئيسياً في ضمان استقرار محاصيل الأرز. يجعل عدم اليقين هذا من الصعب على المزارعين التخطيط لوقت الزراعة الأمثل وإدارة الأراضي. لذلك، هناك حاجة إلى نهج مبتكر وقائم على التكنولوجيا للتنبؤ بغلة المحاصيل بشكل أكثر دقة. تهدف هذه الدراسة إلى بناء نموذج تنبؤي لمحصول الأرز باستخدام الشبكة العصبية الاصطناعية (ANN)، وخاصة بنية *Perceptron* متعدد الطبقات (MLP)، والتي تكون قادرة على نمذجة العلاقة غير الخطية بين العوامل البيئية وإنتاج الأرز. تم تصميم نموذج ANN من خلال دمج متغيرات الطقس مثل هطول الأمطار ودرجة حرارة الهواء والرطوبة، بالإضافة إلى ظروف التربة مثل درجة الحموضة ورطوبة التربة ومحتوى المغذيات كمدخلات رئيسية. تم الحصول على البيانات من وكالة الأرصاد الجوية وعلم المناخ والجيوفيزياء (BMKG) ووكالة الإحصاء المركزية (BPS) لمنطقة مالانج في الفترة من 2009 إلى 2024. تتضمن عملية المعالجة المسبقة للبيانات تطبيع البيانات باستخدام طريقة *Z-Score* بالإضافة إلى معالجة القيمة المفقودة لضمان جودة الإدخال المثلى. بالإضافة إلى ذلك، يتم تطبيق عملية ضبط المعلمة الفائقة باستخدام طريقة *Grid Search* لتحسين أداء النموذج واستقراره، بالإضافة إلى تقليل مخاطر الإفراط في التجهيز وعدم التجهيز. تم إجراء التقييم باستخدام مقاييس متوسط الخطأ المطلق (MAE)، وجذر متوسط الخطأ التربيعي (RMSE)، و  $R^2$  (R-Squared). ومن المتوقع ألا توفر النتائج التي تم الحصول عليها تنبؤات دقيقة فحسب، بل أن تستخدم أيضاً كأساس لاتخاذ القرارات في تخطيط الأمن الغذائي التكيفي والقائم على البيانات. وبالتالي، يساهم هذا البحث في تطوير الزراعة الدقيقة في إندونيسيا التي تكون أكثر حداثة وكفاءة واستدامة.

**الكلمات المفتاحية:** الشبكة العصبية الاصطناعية، التنبؤ بالعائد، ضبط المعلمة الفائقة

## DAFTAR ISI

COVER .....	i
PERNYATAAN KEASLIAN TULISAN .....	iv
HALAMAN PERSETUJUAN UJIAN THESIS .....	v
HALAMAN PENGESAHAN .....	vi
ABSTRAK .....	viii
ABSTRACT .....	ix
المخلص .....	x
DAFTAR ISI .....	xi
DAFTAR GAMBAR .....	xiii
DAFTAR TABEL .....	xiv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	8
1.3 Tujuan Penelitian .....	8
1.4 Riset Objek .....	8
1.5 Batasan Masalah .....	9
1.6 Manfaat Penelitian .....	10
BAB II STUDI PUSTAKA .....	11
2.1 Landasan teori .....	11
2.1.1 <i>Machine learning</i> .....	11
2.1.2 <i>Z-Score</i> .....	12
2.2 Penelitian Terdahulu .....	20
BAB III METODE PENELITIAN .....	26

3.1	Alur Penelitian .....	26
3.2	Desain Penelitian.....	32
3.2.1	Pengumpulan dan Pemahaman Data.....	33
3.2.2	Data <i>preprocessing</i> .....	34
3.2.3	Feature Selection.....	37
3.2.4	Modelling menggunakan Algoritma <i>ANN</i> .....	39
3.2.5	Optimasi model ( <i>Hyperparameter tuning</i> ) .....	53
BAB IV HASIL DAN PEMBAHASAN.....		63
4.1.	Hasil Pengumpulan dan Pemahaman Data .....	63
4.1.1.	Variabel yang digunakan .....	63
4.1.2.	Distribusi dan Statistik Data .....	65
4.2.	Hasil <i>Preprocessing Data</i> .....	70
4.3.	Modelling dengan <i>ANN</i> .....	79
4.4.	Hyperparameter Tuning dengan Gridsearch .....	83
4.5.	Korelasi Hasil Penelitian dalam Perspektif Islam.....	92
BAB V.....		98
KESIMPULAN .....		98
5.1.	Kesimpulan .....	98
5.2.	Saran.....	99
DAFTAR PUSTAKA .....		99

## DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian .....	26
Gambar 3.2 Arsitektur <i>ANN</i> pada Prediksi Panen Padi.....	41
Gambar 4.1 Visualisasi Grafik batang produksi pada tiap kecamatan.....	70
Gambar 4.2 Hasil visualisasi kurva <i>training ANN</i> .....	84
Gambar 4.3 Visualisasi hasil <i>Actual vs Predicted ANN</i> .....	85
Gambar 4.4 Visualisasi prediksi dan aktual <i>ANN</i> setelah <i>Grid Search</i> .....	93
Gambar 4.5 (a) Visualisasi Hasil MAE (b) MSE (c) $R^2$ -Score.....	95

## DAFTAR TABEL

Tabel 2.1 Daftar Jurnal Relevan .....	24
Tabel 3.1 Variabel Terikat dan Bebas Penelitian.....	39
Tabel 3.2 <i>Pseudocode</i> Gridsearch dalam menentukan parameter .....	58
Tabel 4.1 Variabel yang digunakan .....	65
Tabel 4.2 Statistik deskriptif masing-masing variabel.....	68
Tabel 4.3 Proses Pembersihan Data .....	74
Tabel 4.4 Hasil Ekstraksi Fitur .....	76
Tabel 4.5 Hasil <i>One-Hot Encoding</i> untuk Variabel Kecamatan .....	78
Tabel 4.6 Normalisasi <i>Z-Score</i> .....	81
Tabel 4.7 Hasil Modelling beberapa Algoritma.....	84
Tabel 4.8 Hasil Hyperparameter tuning model .....	91

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Indonesia, sebagai salah satu negara agraris terbesar di dunia, memiliki peran penting dalam penyediaan pangan, terutama padi, yang menjadi sumber makanan utama bagi lebih dari 270 juta penduduknya . Menurut Park, (2020) sebagai negara yang memiliki beragam iklim dan kondisi geografis, Indonesia menghadapi tantangan yang signifikan dalam produksi padi. Para petani sering kali berjuang untuk mencapai hasil panen yang optimal, terutama di tengah ketidakpastian yang disebabkan oleh perubahan iklim dan kondisi tanah yang bervariasi (Dhamira and Irham, 2020). Hal ini menuntut perhatian serius dari berbagai pihak, termasuk pemerintah, peneliti, dan praktisi pertanian, untuk mencari solusi yang efektif demi menjaga ketahanan pangan nasional. Seperti yang di sebutkan dalam firman Allah SWT dalam

QS. Al-An'am (6): 99

وَهُوَ الَّذِي أَنْزَلَ مِنَ السَّمَاءِ مَاءً فَأَخْرَجْنَا بِهِ نَبَاتَ كُلِّ شَيْءٍ فَأَخْرَجْنَا مِنْهُ خَضِرًا  
تُخْرِجُ مِنْهُ حَبًّا مُتَرَاكِبًا وَمِنَ النَّخْلِ مِنْ طَلْعِهَا قِنْوَانٌ دَانِيَةٌ وَجَنَّاتٍ مِنْ أَعْنَابٍ  
وَالزَّيْتُونِ وَالرُّمَّانِ مُشْتَبِهًا وَغَيْرَ مُتَشَابِهٍ انْظُرُوا إِلَى ثَمَرِهِ إِذَا أَثْمَرَ وَيَنْعِهِ إِنَّ فِي ذَلِكَُمْ  
لَآيَاتٍ لِقَوْمٍ يُؤْمِنُونَ

*Artinya :*

*“ Dialah yang menurunkan air dari langit lalu dengannya Kami menumbuhkan segala macam tumbuhan. Maka, darinya Kami mengeluarkan tanaman yang menghijau. Darinya Kami mengeluarkan butir yang bertumpuk (banyak). Dari mayang kurma (mengurai) tangkai-tangkai yang menjuntai kebun-kebun anggur. (Kami menumbuhkan pula) zaitun dan delima yang serupa dan yang tidak serupa. Perhatikanlah buahnya pada waktu berbuah dan menjadi masak. Sesungguhnya pada yang demikian itu benar-benar terdapat tanda-tanda bagi kaum yang beriman. ”*

Ayat ini menggambarkan kebesaran Allah SWT melalui proses alami yang kompleks, yakni turunnya hujan sebagai sumber kehidupan tumbuhan yang menghasilkan berbagai jenis biji-bijian dan buah-buahan. Ibnu Katsir menekankan kekuasaan Allah SWT dalam mengatur pertumbuhan tanaman hingga menghasilkan panen yang siap dikonsumsi, sebagai tanda yang patut direnungi oleh orang beriman (Ramadhan, Septiawardi, and Hendro, 2024).

Tafsir Ibnu Abbas mengungkapkan bahwa meskipun buah-buahan seperti kurma, anggur, dan delima tampak serupa dalam beberapa aspek (misalnya bentuk dan warna), mereka memiliki rasa dan manfaat yang sangat berbeda. Keberagaman ini adalah bukti dari kebijaksanaan Allah SWT dalam menciptakan berbagai jenis tanaman dan buah-buahan yang bermanfaat untuk kehidupan manusia, meskipun berasal dari sumber yang serupa.

Dalam konteks ini, keteraturan dan keberagaman yang ditunjukkan oleh ciptaan Allah SWT menjadi dasar yang sangat relevan bagi pemahaman ilmu pengetahuan modern, khususnya dalam penerapan kecerdasan buatan. Tafsir Al-Jalalayn menyoroti keajaiban alam yang menghasilkan hasil bumi yang serupa namun berbeda, yang juga dapat dipahami sebagai cerminan dari hukum alam yang sangat teratur dan harmonis.



Keteraturan ini, yang telah Allah SWT tetapkan dalam ciptaan-Nya, memberikan inspirasi bagi penerapan teknologi seperti *Artificial Neural Networks (ANN)*, yang memanfaatkan variabel alami seperti curah hujan, suhu, kelembapan, dan kondisi tanah untuk meniru pola-pola alami dalam pertumbuhan tanaman (Qodri and Kamil 2023).

*ANN* berfungsi untuk meniru keteraturan ciptaan Allah SWT , dengan tujuan memprediksi hasil panen secara lebih akurat, seperti halnya Allah SWT mengatur pertumbuhan tanaman dengan cara yang sempurna. Dengan menggunakan data alam ini, manusia dapat mengelola pertanian secara lebih efektif, yang tidak hanya berdampak pada peningkatan hasil panen tetapi juga mencerminkan upaya untuk memahami dan mengikuti sunnatullah dalam ciptaan-Nya. Melalui teknologi, kita dapat lebih menghargai dan memanfaatkan ciptaan Allah SWT dengan cara yang lebih ilmiah, sejalan dengan kebijaksanaan-Nya dalam menciptakan alam semesta yang begitu teratur dan penuh dengan keberagaman yang bermanfaat.

Salah satu tantangan utama dalam produksi padi adalah fluktuasi cuaca yang semakin tidak menentu. Perubahan pola curah hujan dan suhu ekstrem dapat berdampak langsung pada produktivitas tanaman padi (Lobell *et al.*, 2011). Hal ini mengingatkan pada kisah Nabi Yusuf dalam Al-Qur'an, di mana pentingnya mempersiapkan hasil panen berdasarkan prediksi:

*QS. Yusuf [12]: 47-49*

قَالَ تَزْرَعُونَ سَبْعَ سِنِينَ دَابًّا فَمَا حَصَدْتُمْ فَذَرُوهُ فِي سُنْبُلِهِ إِلَّا قَلِيلًا مِّمَّا تَأْكُلُونَ ثُمَّ  
يَأْتِي مِنْ بَعْدِ ذَلِكَ سَبْعٌ شِدَادٌ يَأْكُلْنَ مَا قَدَّمْتُمْ لَهُنَّ إِلَّا قَلِيلًا مِّمَّا تُحْصِنُونَ ثُمَّ يَأْتِي مِنْ  
بَعْدِ ذَلِكَ عَامٌ فِيهِ يُغَاثُ النَّاسُ وَفِيهِ يَعْرِصُونَ<sup>٤</sup>

Artinya :

*"(Yusuf) berkata: 'Kalian akan bercocok tanam selama tujuh tahun berturut-turut; maka apa yang kamu tuai, hendaklah kamu biarkan di bulirnya, kecuali sedikit untuk kamu makan. Kemudian setelah itu akan datang tujuh tahun yang sulit, yang menghabiskan apa yang kalian simpan untuk menghadapi masa sulit itu, kecuali sedikit dari yang kalian simpan. Kemudian setelah itu akan datang tahun di mana manusia diberi hujan (berlimpah) dan pada masa itu mereka memeras (anggur)."*

Ayat QS. Yusuf (12): 47–49 menggambarkan strategi ketahanan pangan dan manajemen krisis yang diilhamkan Allah SWT kepada Nabi Yusuf ‘alaihis-salam saat menafsirkan mimpi Raja Mesir tentang tujuh tahun masa subur dan tujuh tahun masa paceklik, serta masa kemakmuran setelahnya. Tafsir Ibnu Katsir menjelaskan bahwa Yusuf menganjurkan penanaman berkelanjutan dan penyimpanan hasil panen dalam bentuk bulir agar tahan lama, sebagai bentuk perencanaan jangka panjang menghadapi krisis. Tafsir Al-Jalalayn menekankan urutan waktu dan pentingnya kesiapan menghadapi masa sulit dengan strategi penyimpanan yang tepat. Keseluruhan tafsir ini menunjukkan bahwa ayat tersebut bukan sekadar kisah sejarah, melainkan mengandung prinsip-prinsip penting dalam pengelolaan pertanian berkelanjutan, penyimpanan pangan, dan kesiapan menghadapi bencana, yang sangat relevan dengan tantangan era pertanian modern. Dalam konteks ini, penerapan *Artificial Neural*

*Network* (*ANN*) untuk prediksi hasil panen padi menjadi sangat relevan, karena mencerminkan semangat Nabi Yusuf dalam merespons perubahan iklim dan kondisi agrikultur melalui pendekatan prediktif yang berbasis data. *ANN* mampu mengolah data historis seperti cuaca dan kondisi tanah untuk memprediksi hasil panen secara lebih akurat, sehingga membantu petani dan pembuat kebijakan dalam merencanakan produksi, menyimpan hasil panen secara efisien (Amelia *et al.*, 2025).

Nilai-nilai ini sangat relevan dengan tantangan pertanian modern, terutama dalam menghadapi ketidakpastian iklim dan kondisi lahan. Misalnya, Curah hujan yang tidak teratur dapat menyebabkan kekeringan atau banjir yang merusak tanaman. Kondisi ini menciptakan ketidakpastian dalam produksi padi (Saud *et al.*, 2022). Selain itu, kondisi tanah, termasuk tingkat keasaman (pH) dan kelembapan, juga memainkan peran krusial dalam menentukan kualitas serta kuantitas hasil panen (Zhang *et al.*, 2023). Ketidakpastian ini menciptakan kesulitan bagi para petani dalam merencanakan produksi, yang pada akhirnya mengancam ketahanan pangan masyarakat. Oleh karena itu, sangat penting untuk mengembangkan metode yang dapat memprediksi hasil panen dengan akurasi tinggi, sehingga para petani dan pemangku kepentingan lainnya dapat mengambil keputusan yang lebih baik dan lebih tepat waktu.

Di tengah pesatnya perkembangan teknologi, pendekatan *Artificial Intelligence* (AI) mulai diadopsi dalam berbagai sektor, termasuk pertanian. Salah satu metode AI yang menjanjikan dalam memprediksi hasil panen adalah *ANN* (Khairunniza Bejo *et al.*, 2014). *ANN* memiliki kemampuan untuk mengenali pola kompleks antara berbagai variabel yang memengaruhi produksi tanaman. Berbeda dengan metode statistik tradisional yang hanya mampu mengidentifikasi hubungan linear sederhana, *ANN*

dapat memodelkan hubungan non-linear antara variabel-variabel cuaca dan tanah dengan hasil panen (Baral *et al.*, 2011). Kemampuan ini memungkinkan *ANN* untuk menghasilkan prediksi yang lebih akurat, terutama ketika data yang tersedia bersifat tidak teratur dan sulit diprediksi (Eli-Chukwu, 2019).

Beberapa penelitian sebelumnya telah mengeksplorasi model *deep learning* lain dalam prediksi hasil panen, seperti *Long Short-Term Memory* (LSTM), yang memiliki kemampuan dalam menangani data deret waktu. Misalnya, penelitian oleh Bhimavarapu *et al.* (2023) mengembangkan model LSTM yang dioptimalkan dengan *Improved Optimization Function* (IOF) dan membandingkannya dengan metode lain, seperti *Convolutional Neural Network* (CNN), *Recurrent Neural Network* (RNN), dan *Gated Recurrent Unit* (GRU). Hasil penelitian mereka menunjukkan bahwa model IOF-LSTM mampu mengurangi error dalam prediksi hasil panen dengan RMSE sebesar 2.19 dan MAE sebesar 25.4. Sebagian besar studi sebelumnya lebih banyak berfokus pada model berbasis LSTM dan CNN, tetapi masih terbatas dalam mengeksplorasi *ANN* untuk mengatasi ketidakpastian prediksi hasil panen berbasis faktor cuaca dan tanah.

Oleh karena itu, penelitian ini akan berfokus sepenuhnya pada *ANN* sebagai metode utama dalam prediksi hasil panen padi. Penelitian ini akan menggunakan *ANN* dalam seluruh tahapan, termasuk optimasi *hyperparameter*, untuk mendapatkan model yang lebih akurat dan stabil. Meskipun *ANN* menawarkan banyak keunggulan, seperti kemampuan memodelkan hubungan kompleks dan adaptasi terhadap data besar, metode ini juga memiliki kelemahan. Keterbatasan dalam interpretabilitas model, ketergantungan pada data berkualitas tinggi, dan kebutuhan komputasi yang tinggi

menjadi tantangan tersendiri (Albahar, 2023; Ribeiro, Singh, and Guestrin 2016). Oleh karena itu, penelitian ini berfokus pada pengembangan model *ANN* yang lebih komprehensif dengan mengintegrasikan data cuaca dan kondisi tanah sebagai variabel *input*. Dengan pendekatan ini, diharapkan dapat memberikan hasil prediksi yang lebih realistis dan akurat sesuai dengan kondisi nyata yang dihadapi di lapangan (Eli-Chukwu, 2019b).

Selain itu, penelitian ini juga akan melakukan tuning *hyperparameter* pada *ANN* untuk meningkatkan akurasi dan stabilitas model. Dalam penelitian terdahulu oleh (Rosa *et al.*, 2020) *ANN* sering kali mengalami permasalahan *overfitting* atau *underfitting* yang dapat mengurangi keakuratan prediksi. Oleh karena itu, optimasi *hyperparameter* akan diterapkan untuk memastikan bahwa *ANN* dapat memberikan performa yang lebih baik dalam memprediksi hasil panen padi secara akurat.

Kebaruan dari penelitian ini terletak pada pemanfaatan *ANN* dalam memprediksi hasil panen padi dengan mempertimbangkan faktor cuaca dan kondisi tanah secara simultan. Sebagian besar penelitian sebelumnya telah mengeksplorasi metode LSTM dalam prediksi hasil panen, tetapi masih terdapat gap dalam penerapan *ANN* untuk skenario yang lebih kompleks dalam pertanian. Dengan fokus pada *ANN*, penelitian ini berusaha mengevaluasi efektivitas model *ANN* dibandingkan metode yang telah digunakan dalam penelitian sebelumnya, serta meningkatkan interpretabilitas model prediksi bagi para pemangku kepentingan di sektor pertanian.

Dengan demikian, penelitian ini memiliki potensi besar untuk menjadi referensi dalam pengembangan teknologi prediksi hasil panen yang lebih canggih dan aplikatif di masa depan. Diharapkan, hasil penelitian ini dapat memberikan dampak positif bagi

kesejahteraan masyarakat petani dan ketahanan pangan nasional, serta mendorong inovasi lebih lanjut dalam bidang pertanian berkelanjutan.

## 1.2 Rumusan Masalah

1. Bagaimana performa *Artificial Neural Network (ANN)* dalam memprediksi hasil panen padi berdasarkan faktor cuaca dan kondisi tanah?
2. Bagaimana pengaruh tuning *hyperparameter* terhadap performa model *ANN* dalam prediksi hasil panen padi?

## 1.3 Tujuan Penelitian

1. Menganalisis performa *ANN* dalam memprediksi hasil panen padi dengan mengukur akurasi model menggunakan metrik evaluasi seperti *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *R-Squared ( $R^2$ )*.
2. Mengoptimalkan *hyperparameter ANN* untuk meningkatkan akurasi prediksi hasil panen padi dan mengurangi kesalahan prediksi.

## 1.4 Riset Objek

Objek penelitian ini adalah model prediksi hasil panen padi yang dikembangkan menggunakan pendekatan (*ANN*) dengan memanfaatkan data variabel cuaca dan kondisi tanah sebagai *input* utama. Penelitian ini berfokus pada pengembangan dan evaluasi akurasi model *ANN* dalam memprediksi hasil panen padi di berbagai wilayah pertanian di Kota Malang selama rentang tahun 2009–2024, berdasarkan data historis cuaca dan kualitas tanah. Model yang dikembangkan diharapkan dapat menjadi alat

bantu bagi petani dan pemangku kepentingan dalam merencanakan produksi pertanian secara lebih optimal melalui prediksi hasil panen yang lebih akurat.

### **1.5 Batasan Masalah**

Penelitian ini dibatasi pada beberapa aspek untuk memastikan fokus dan kedalaman analisis, yaitu:

#### **1. Variabel yang Digunakan Terbatas**

Penelitian ini hanya mempertimbangkan faktor cuaca (seperti suhu, curah hujan, kelembaban, dan lainnya) serta luas panen sebagai variabel input dalam pemodelan prediksi hasil panen padi.

#### **2. Tidak Melibatkan Faktor Agronomis Lain**

Variabel agronomis penting seperti jenis atau varietas padi, penggunaan pupuk, teknik irigasi, frekuensi tanam, serta pengendalian hama dan penyakit tidak dimasukkan dalam analisis.

#### **3. Tidak Mencakup Aspek Sosial-Ekonomi**

Penelitian ini belum mempertimbangkan faktor sosial-ekonomi petani seperti tingkat pendidikan, pendapatan, kepemilikan lahan, atau akses terhadap teknologi pertanian yang juga dapat mempengaruhi hasil panen.

#### **4. Keterbatasan Wilayah Studi**

Penelitian ini difokuskan pada wilayah tertentu saja, sehingga hasil dan model yang dikembangkan belum tentu dapat digeneralisasikan ke seluruh wilayah pertanian di Indonesia.

## 1.6 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat, baik bagi dunia akademis maupun praktisi di sektor pertanian:

1. Manfaat bagi petani. Memberikan informasi yang lebih akurat mengenai prediksi hasil panen, sehingga petani dapat melakukan perencanaan pola tanam yang lebih tepat dan mengantisipasi risiko gagal panen akibat perubahan cuaca atau kondisi tanah yang kurang mendukung.
2. Manfaat bagi pemerintah dan pembuat kebijakan. Menyediakan data prediktif yang dapat digunakan untuk perencanaan ketahanan pangan nasional, terutama dalam menghadapi perubahan iklim dan menjaga stabilitas produksi padi.
3. Manfaat bagi akademisi dan peneliti. Menyumbangkan literatur baru dalam penggunaan *Artificial Neural Network* untuk prediksi hasil panen, khususnya dalam komoditas padi, yang dapat menjadi dasar penelitian lanjutan dalam bidang pertanian berbasis teknologi.



## BAB II

### STUDI PUSTAKA

#### 2.1 Landasan teori

##### 2.1.1 *Machine learning*

*Machine learning* (ML) merupakan cabang dari kecerdasan buatan (*Artificial Intelligence - AI*) yang memungkinkan sistem komputer untuk belajar dari data tanpa diprogram secara eksplisit (Kühl *et al.*, 2022). Teknologi ini mengandalkan algoritma yang mampu mengenali pola dalam data dan menggeneralisasi pola tersebut untuk membuat prediksi atau keputusan. Dalam bidang pertanian, khususnya dalam prediksi hasil panen, ML digunakan untuk menganalisis berbagai faktor seperti kondisi cuaca, kelembaban tanah, suhu, dan curah hujan guna menghasilkan estimasi produktivitas tanaman dengan tingkat akurasi yang tinggi (Sarker, 2021).

*Machine learning* dapat diklasifikasikan menjadi tiga jenis utama, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. *Supervised learning* adalah pendekatan yang paling umum digunakan dalam prediksi hasil panen, di mana model dilatih menggunakan dataset yang memiliki label atau *output* yang sudah diketahui (Mich, 2020). Model akan belajar dari pasangan *input-output* dan menemukan pola yang dapat digunakan untuk membuat prediksi pada data baru. Beberapa algoritma yang sering digunakan dalam *supervised learning* meliputi regresi linear, regresi logistik, *support vector machine* (SVM), *decision tree*, *random forest*, dan *Artificial Neural Network* (ANN).

label, sehingga algoritma berfokus pada pengelompokan atau menemukan struktur tersembunyi dalam data. Contoh algoritma yang digunakan dalam *unsupervised learning* antara lain *k-means clustering* dan *hierarchical clustering*, yang sering diterapkan dalam segmentasi data atau identifikasi pola tersembunyi. Sementara itu, *reinforcement learning* adalah metode yang memungkinkan model belajar melalui interaksi dengan lingkungan dan menerima umpan balik dalam bentuk reward atau penalti untuk mengoptimalkan keputusan yang diambil. Pendekatan ini sering digunakan dalam sistem kontrol, robotika, dan optimasi keputusan. *Machine learning* telah menjadi teknologi yang sangat bermanfaat dalam berbagai bidang, termasuk pertanian, kesehatan, keuangan, dan industri. Kemampuannya dalam mengolah data dalam jumlah besar serta mengidentifikasi pola yang kompleks menjadikannya salah satu pendekatan yang sangat efektif dalam analisis dan prediksi berbasis data (Rosa *et al.*, 2020).

### 2.1.2 Z-Score

*Z-Score* merupakan salah satu metode normalisasi data yang digunakan untuk mengubah skala fitur ke distribusi dengan rata-rata nol dan standar deviasi satu. Teknik ini sering diterapkan dalam *Machine learning* untuk memastikan bahwa setiap fitur memiliki kontribusi yang seimbang dalam proses pelatihan model, terutama ketika dataset memiliki skala yang berbeda-beda (Singh, 2020).

Normalisasi sangat penting dalam algoritma *Machine learning*, terutama yang berbasis optimasi gradien seperti (*ANN*) dan regresi logistik. Ketika fitur memiliki skala yang sangat berbeda, algoritma optimasi dapat mengalami kesulitan dalam

konvergensi, karena bobot yang diperbarui akan lebih dipengaruhi oleh fitur dengan skala lebih besar. Dengan menggunakan *Z-Score*, setiap fitur akan memiliki distribusi yang lebih seragam, sehingga mempercepat proses pelatihan dan meningkatkan stabilitas model. Adapun untuk rumus *Z-Score* dijelaskan pada Persamaan 2.1 (Herwanto *et al.*, 2021).

$$Z = \frac{X - \mu}{\sigma} \quad (2.1)$$

Keterangan:

$X$  adalah nilai asli dari fitur,

$\mu$  adalah rata-rata dari fitur tersebut,

$\sigma$  adalah standar deviasi dari fitur tersebut.

Hasil dari transformasi *Z-Score* akan menghasilkan distribusi data dengan mean (rata-rata) sebesar 0 dan standar deviasi sebesar 1, sehingga data memiliki skala yang lebih seragam.

Salah satu keuntungan utama dari normalisasi *Z-Score* adalah kemampuannya dalam mempertahankan distribusi asli data, sehingga tetap efektif meskipun ada nilai *outlier* dalam dataset. Berbeda dengan metode normalisasi lain seperti *Min-Max Scaling* yang dapat terpengaruh oleh nilai ekstrim, *Z-Score* tetap menjaga struktur distribusi data secara lebih stabil (Singh, 2022).

Dalam penerapan *Machine learning*, normalisasi *Z-Score* sering digunakan dalam tahap *preprocessing* data sebelum model dilatih. Setelah data dinormalisasi, model akan lebih mudah memahami hubungan antar variabel tanpa dipengaruhi oleh

perbedaan skala. Oleh karena itu, teknik ini menjadi salah satu pendekatan yang banyak digunakan dalam berbagai studi pemodelan prediktif, termasuk dalam analisis dan prediksi hasil panen berbasis *Machine learning*.

### **2.1.3 Artificial Neural Network**

Jaringan Saraf Tiruan (*Artificial Neural Network* - *ANN*) merupakan salah satu metode dalam *Machine learning* yang terinspirasi dari cara kerja otak manusia dalam memproses informasi (Joshi *et al.*, 2021). *ANN* terdiri dari kumpulan *neuron* buatan yang terhubung dalam beberapa lapisan dan bekerja secara bersamaan untuk mengenali pola dalam data. Model ini sangat efektif dalam menangani masalah kompleks, terutama dalam pemodelan non-linear dan pemrosesan data dalam jumlah besar. *ANN* banyak diterapkan dalam berbagai bidang, termasuk pengenalan pola, pengolahan citra, prediksi keuangan, dan pertanian, termasuk dalam estimasi hasil panen berdasarkan faktor lingkungan dan tanah.

Secara umum, *ANN* terdiri dari tiga jenis lapisan utama, yaitu lapisan *input* (*input layer*), lapisan tersembunyi (*Hidden Layer*), dan lapisan *output* (*output layer*). Lapisan *input* menerima data mentah yang akan diproses lebih lanjut oleh lapisan tersembunyi melalui bobot dan bias yang diperbarui selama proses pembelajaran. Lapisan tersembunyi bertanggung jawab atas ekstraksi fitur dan representasi data yang lebih abstrak, sementara lapisan *output* menghasilkan prediksi akhir sesuai dengan tujuan pemodelan. Jumlah *neuron* dan lapisan tersembunyi sangat memengaruhi kompleksitas serta kemampuan model dalam menangkap pola dalam data (Abdolrasol *et al.*, 2021).

Jaringan Saraf Tiruan terus berkembang dengan adanya inovasi dalam arsitektur dan teknik pelatihan. Beberapa varian *ANN* yang lebih kompleks, seperti *Deep Neural Networks* (DNN), *Convolutional Neural Networks* (CNN), dan *Recurrent Neural Networks* (RNN), telah banyak digunakan dalam berbagai aplikasi canggih, termasuk pengenalan wajah, pemrosesan bahasa alami (NLP), serta prediksi berbasis data besar. Dengan fleksibilitas dan kemampuannya dalam menangani data non-linear, *ANN* menjadi salah satu pendekatan paling efektif dalam berbagai aplikasi kecerdasan buatan (Montesinos López *et al.*, 2022).

#### **2.1.4 Hyperparameter tuning**

*Hyperparameter tuning* adalah proses optimasi nilai *hyperparameter* dalam model *Machine learning* untuk meningkatkan kinerja dan generalisasi model terhadap data baru (Yang and Shami, 2020). *Hyperparameter* adalah parameter yang nilainya ditentukan sebelum proses pelatihan model dimulai dan tidak diperbarui selama proses pembelajaran. Pemilihan *hyperparameter* yang optimal sangat penting karena dapat mempengaruhi kecepatan konvergensi, akurasi model, serta kemampuannya dalam menghindari *overfitting* atau *underfitting*. Beberapa *hyperparameter* yang sering disesuaikan dalam berbagai algoritma *Machine learning* meliputi laju pembelajaran (*Learning rate*), jumlah iterasi atau *Epoch*, jumlah dan ukuran batch data yang digunakan dalam pelatihan, serta teknik regularisasi untuk mengurangi kompleksitas model.

Terdapat berbagai metode yang digunakan dalam *hyperparameter tuning* untuk menemukan kombinasi terbaik. Salah satu metode yang paling umum adalah *Grid*

*Search*, yang melakukan pencarian sistematis dengan mencoba setiap kombinasi *hyperparameter* dalam ruang pencarian yang telah ditentukan. Meskipun metode ini menjamin bahwa semua kemungkinan diuji, *Grid Search* sering kali memerlukan sumber daya komputasi yang besar, terutama jika jumlah *hyperparameter* yang diuji cukup banyak. Alternatif yang lebih efisien adalah Random Search, yang memilih kombinasi *hyperparameter* secara acak dalam ruang pencarian (Liao *et al.*, (2022)). Meskipun tidak menjamin menemukan kombinasi optimal, metode ini sering kali lebih efisien dibandingkan *Grid Search* karena dapat mengeksplorasi berbagai kemungkinan dengan jumlah iterasi yang lebih sedikit.

Pendekatan lain yang lebih canggih adalah *Bayesian Optimization*, yang menggunakan model probabilistik untuk memperkirakan kombinasi *hyperparameter* terbaik berdasarkan hasil dari eksperimen sebelumnya. Dengan cara ini, *Bayesian Optimization* dapat mengurangi jumlah eksperimen yang diperlukan untuk menemukan konfigurasi optimal, sehingga lebih efisien dibandingkan metode pencarian berbasis grid atau acak. Selain itu, terdapat metode *Hyperband*, yang menggunakan pendekatan adaptif dengan mengalokasikan sumber daya lebih efisien berdasarkan performa awal model. Jika suatu kombinasi *hyperparameter* tidak menunjukkan hasil yang menjanjikan dalam tahap awal pelatihan, metode ini dapat menghentikan eksperimen lebih awal untuk menghemat sumber daya (Bischl *et al.*, 2023).

Beberapa algoritma optimasi berbasis evolusi, seperti *Genetic Algorithm* (GA) dan *Evolutionary Strategies*, juga digunakan dalam *hyperparameter tuning* dengan prinsip seleksi alam. Metode ini memodifikasi nilai *hyperparameter* dari generasi ke generasi untuk menemukan kombinasi yang optimal. Meskipun membutuhkan waktu

yang lebih lama, pendekatan ini efektif dalam mengeksplorasi ruang pencarian yang luas dan dapat menemukan solusi optimal dalam model yang kompleks.

*Hyperparameter tuning* menjadi langkah krusial dalam *Machine learning* karena dapat secara signifikan meningkatkan performa model (Shahhosseini *et al.*, 2022). Dengan memilih metode tuning yang sesuai dengan kompleksitas model dan ketersediaan sumber daya komputasi, proses pencarian *hyperparameter* optimal dapat dilakukan dengan lebih efisien dan menghasilkan model yang lebih akurat serta dapat diandalkan dalam melakukan prediksi atau pengambilan keputusan berbasis data.

### 2.1.5 Mean Absolute Error (MAE)

*Mean Absolute Error* (MAE) adalah salah satu metrik evaluasi dalam *Machine learning* yang digunakan untuk mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual (Karunasingha, 2022a). MAE menunjukkan seberapa besar rata-rata kesalahan yang dibuat oleh model dalam memprediksi data. Semakin kecil nilai MAE, semakin baik performa model dalam menghasilkan prediksi yang mendekati nilai sebenarnya. Adapun Rumus MAE dapat dituliskan seperti pada Persamaan 2.2 (Karunasingha, 2022).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.2)$$

Keterangan:

$n$  adalah jumlah sampel dalam dataset,

$y_i$  adalah nilai aktual,

$\hat{y}_i$  adalah nilai prediksi,

$|y_i - \hat{y}_i|$  adalah selisih absolut antara nilai aktual dan nilai prediksi.

MAE memiliki sifat yang mudah diinterpretasikan karena satuannya sama dengan satuan data aslinya. Selain itu, metrik ini tidak terlalu sensitif terhadap *outlier* dibandingkan dengan metrik lain seperti *Root Mean Squared Error* (RMSE). Namun, karena MAE tidak melakukan kuadrat terhadap kesalahan, model yang menggunakan MAE sebagai metrik evaluasi tidak akan memberikan penalti yang lebih besar terhadap kesalahan besar, sehingga dalam beberapa kasus, RMSE lebih dipilih ketika kesalahan besar perlu mendapat perhatian lebih (Qi *et al.*, 2020).

Dalam penerapan *Machine learning*, MAE sering digunakan dalam berbagai masalah regresi, termasuk prediksi harga, perkiraan cuaca, dan estimasi hasil panen. Nilai MAE yang lebih kecil menunjukkan bahwa model memiliki tingkat akurasi yang lebih baik dalam memprediksi nilai target, sementara nilai MAE yang besar menunjukkan bahwa model masih memiliki tingkat kesalahan yang tinggi dan perlu dilakukan optimasi lebih lanjut.

#### **2.1.6 Root Mean Squared Error (RMSE)**

*Root Mean Squared Error* (RMSE) adalah metrik evaluasi dalam *Machine learning* yang digunakan untuk mengukur rata-rata kesalahan kuadrat antara nilai prediksi dan nilai aktual. RMSE memberikan gambaran seberapa jauh prediksi model menyimpang dari nilai sebenarnya dengan memberikan penalti lebih besar pada kesalahan yang lebih besar. Rumus RMSE dapat dituliskan seperti pada Persamaan 2.3 (Plevris *et al.*, 2022).



$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.3)$$

Keterangan:

$n$  adalah jumlah data,

$y_i$  adalah nilai aktual,

$\hat{y}_i$  adalah nilai prediksi,

$(y_i - \hat{y}_i)^2$  adalah selisih kuadrat antara nilai aktual dan nilai prediksi.

RMSE dihitung dengan mengambil akar dari rata-rata jumlah kesalahan kuadrat, sehingga nilai yang diperoleh berada dalam satuan yang sama dengan variabel target yang diprediksi. Metrik ini sering digunakan dalam berbagai aplikasi *Machine learning*, terutama dalam masalah regresi, untuk mengukur tingkat kesalahan model terhadap data yang digunakan (Kumar *et al.*, 2022).

Nilai RMSE yang lebih rendah menunjukkan bahwa model memiliki tingkat kesalahan yang lebih kecil dalam melakukan prediksi, sedangkan nilai yang lebih tinggi menunjukkan adanya penyimpangan yang lebih besar antara nilai prediksi dan nilai aktual (Ağbulut *et al.*, 2021). Oleh karena itu, RMSE dapat digunakan untuk mengevaluasi performa model dan membandingkan berbagai algoritma dalam mencari model yang paling akurat.

### 2.1.6 *R-Squared* ( $R^2$ )

*R-Squared* ( $R^2$ ) adalah metrik evaluasi yang digunakan untuk mengukur seberapa baik model regresi dapat menjelaskan variabilitas dalam data target. Nilai  $R^2$  menunjukkan proporsi variasi dalam data yang dapat dijelaskan oleh model, dengan

rentang nilai antara 0 hingga 1. Semakin tinggi nilai  $R^2$ , semakin baik model dalam menjelaskan hubungan antara variabel independen dan variabel dependen . Adapun Rumus  $R^2$  dapat dituliskan seperti Persamaan 2.4 (Rights and Sterba, 2020)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.4)$$

Keterangan :

$y_i$  adalah nilai aktual,

$\hat{y}_i$  adalah nilai prediksi,

$\bar{y}$  adalah rata-rata dari nilai aktual,

$\sum (y_i - \hat{y}_i)^2$  adalah jumlah kuadrat kesalahan (Residual Sum of Squares, RSS),

$\sum (y_i - \bar{y})^2$  adalah jumlah kuadrat total (Total Sum of Squares, TSS).

Nilai  $R^2$  yang mendekati 1 menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam menjelaskan variabilitas data, sedangkan nilai yang mendekati 0 menunjukkan bahwa model kurang mampu menjelaskan hubungan antara variabel prediktor dan target (Tyagi *et al.*, 2022). Dalam beberapa kasus, nilai  $R^2$  bisa bernilai negatif, yang menunjukkan bahwa model tidak lebih baik dari sekadar menggunakan rata-rata sebagai prediksi. Sebagai metrik evaluasi,  $R^2$  sering digunakan dalam analisis regresi untuk menilai sejauh mana model dapat memprediksi hasil dengan akurat berdasarkan pola dalam data yang digunakan (Sudin *et al.*, 2021).

## 2.2 Penelitian Terdahulu

Penelitian oleh Kujawa dan Niedbała (2021) dalam jurnal *Agriculture* menunjukkan bahwa jaringan saraf tiruan (*ANN*) dapat meningkatkan efisiensi dan kualitas produksi pertanian melalui penerapannya dalam berbagai aspek, seperti prediksi hasil panen, diagnosis penyakit tanaman, serta optimalisasi manajemen

pertanian, dengan mempertimbangkan variabel independen seperti faktor cuaca dan kondisi tanah. Penelitian ini juga menekankan pentingnya integrasi pembelajaran mesin dalam pertanian presisi dan digital untuk mendukung peningkatan produktivitas secara berkelanjutan.

Sementara itu, Agussabti *et al.* (2022) dalam jurnal *Open Agriculture* menyoroti tantangan adopsi teknologi pertanian pintar (*Smart Farming Technology/SFT*) oleh petani kecil di Aceh, di mana meskipun terdapat persepsi positif terhadap teknologi tersebut, kesiapan adopsinya masih rendah akibat keterbatasan kapasitas teknis, pendidikan, dan akses pembiayaan. Untuk itu, diperlukan program pelatihan serta dukungan finansial yang memadai. Selanjutnya, Ahmed *et al.* (2024) dalam jurnal *IET Smart Cities* mengembangkan sistem klasifikasi biji kopi hijau berbasis *Convolutional Neural Network* (CNN) yang mampu melakukan klasifikasi kualitas secara *real-time* dengan akurasi tinggi dan dapat diimplementasikan langsung di lokasi produksi melalui perangkat *embedded*, sehingga mendukung efisiensi produksi dan pengembangan pertanian cerdas di sektor kopi. Mulyono *et al.* (2024) dalam jurnal *Scientific Reports* memanfaatkan model *ANN* hibrida yang dikombinasikan dengan algoritma optimasi untuk meningkatkan akurasi prediksi suhu tanah di ekosistem padang rumput subtropika Florida, dan hasilnya menunjukkan bahwa model ini tidak hanya lebih akurat dibanding model tradisional, tetapi juga mampu menangkap variasi harian dan musiman suhu tanah secara efektif, yang penting dalam pengelolaan lahan dan adaptasi terhadap perubahan iklim. Terakhir, Sharma *et al.* (2022) membandingkan beberapa algoritma pembelajaran mesin—termasuk *Support Vector Machine* (SVM),

*Decision Tree* (DT), dan *Random Forest* (RF)—dalam mendeteksi penyakit tanaman kapas berbasis citra daun, dan menemukan bahwa RF menghasilkan akurasi tertinggi sebesar 92%, mengindikasikan bahwa metode ini sangat potensial untuk diterapkan dalam sistem deteksi penyakit tanaman secara otomatis dan efisien.

Berdasarkan kajian literatur terdahulu dan hasil penelitian yang ada, seperti yang dilakukan oleh Basir *et al.* (2021) yang menggunakan *ANN* untuk memprediksi hasil panen padi di Bangladesh, model *ANN* terbukti lebih unggul dibandingkan model-model tradisional dalam hal akurasi prediksi dan adaptasi terhadap variabel yang berubah. Hal ini menunjukkan bahwa *ANN* memiliki potensi besar dalam mengatasi tantangan ketidakpastian yang dihadapi dalam produksi pertanian, terutama dalam memprediksi hasil panen yang dipengaruhi oleh fluktuasi cuaca dan kondisi tanah yang tidak dapat diprediksi secara linear.

Secara keseluruhan, kebaruan dari penelitian ini terletak pada penerapan *ANN* sebagai metode utama dalam memprediksi hasil panen padi dengan mempertimbangkan faktor cuaca dan kondisi tanah secara simultan, serta pemanfaatan teknologi *AIoT* untuk mengumpulkan data *real-time*. Pendekatan ini diharapkan dapat menghasilkan model prediksi yang lebih akurat, fleksibel, dan aplikatif dalam konteks pertanian, terutama dalam menghadapi tantangan perubahan iklim dan ketidakpastian kondisi tanah. Kemudian untuk memperjelas penelitian terdahulu, penjelasan mengenai jurnal-jurnal yang relevan dengan penelitian ini dapat disajikan dalam Tabel 2.1.

Tabel 2.1 Jurnal Relevan

No.	Judul Jurnal	Metode Penelitian	Penulis dan Tahun	Hasil Penelitian
1	<i>Performance Evaluation of Machine learning Techniques for Mustard Crop Yield Prediction from Soil Analysis</i>	<i>KNN, Naïve Bayes, Logistic Regression, Artificial Neural Network (ANN), dan Random Forest</i>	Vaishali Pandith <i>et al.</i> (2020)	<i>ANN dan KNN menunjukkan performa tinggi dalam prediksi hasil panen mustard; ANN memiliki nilai f-score tertinggi (0.9976) dan precision mencapai 99.94%, menandakan kemampuannya dalam mengidentifikasi pola dari data tanah dengan sangat akurat.</i>
2	<i>Artificial Neural Network model in predicting yield of mechanically transplanted rice from transplanting parameters in Bangladesh</i>	<i>Artificial Neural Networks (ANN)</i>	Md Samiul Basir <i>et al</i> (2021)	Penelitian ini menunjukkan bahwa model berbasis <i>ANN</i> akan menjadi alternatif model regresi dan metode prediksi hasil panen yang lebih akurat berdasarkan parameter lahan tanam.
3	<i>Using machine learning for crop yield prediction in the past or the future</i>	<i>Regularized Linear Models, RF, ANN</i>	A. Morales, F.J. Villalobos (2023)	<i>ANN</i> menunjukkan performa baik dalam prediksi hasil panen masa kini dan mendatang.
4	<i>Enhancing direct-seeded rice yield prediction using UAV-derived features acquired during the reproductive phase</i>	<i>MLR, PLSR, ANN, RF, SVR</i>	G. Yang <i>et al.</i> (2024)	<i>ANN</i> efektif memanfaatkan fitur UAV untuk prediksi hasil panen padi.

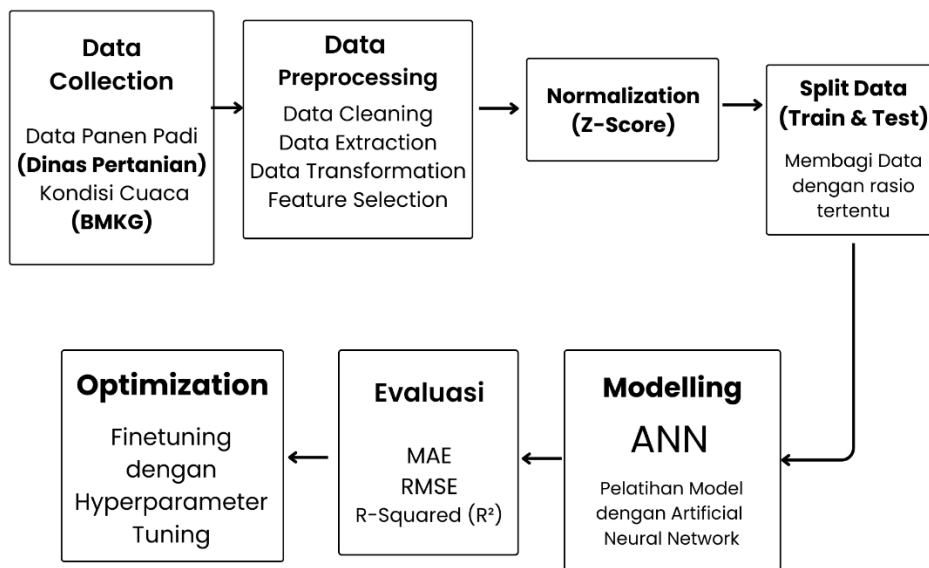
No.	Judul Jurnal	Metode Penelitian	Penulis dan Tahun	Hasil Penelitian
5	<i>Performance Evaluation of Machine learning Techniques for Mustard Crop Yield Prediction from Soil Analysis</i>	<i>KNN, Naïve Bayes, Multinomial Logistic Regression, ANN, RF</i>	V. Pandith <i>et al.</i> (2020)	<i>ANN</i> termasuk metode dengan akurasi tinggi dalam prediksi hasil panen berdasarkan analisis tanah.

## BAB III

### METODE PENELITIAN

#### 3.1 Alur Penelitian

Kerangka konseptual penelitian ini ditunjukkan pada Gambar 3.1, yang memvisualisasikan tahapan penelitian mulai dari pengumpulan data hingga evaluasi model. Alur ini menggambarkan integrasi antara faktor cuaca, kondisi tanah, dan metode (*ANN*) untuk menghasilkan prediksi hasil panen padi yang akurat.



Gambar 3.1 Diagram Alur penelitian

### 3.1.1 Pengumpulan Data

Pengumpulan data merupakan tahap pertama dan paling fundamental dalam penelitian ini karena kualitas data sangat menentukan akurasi model prediksi. Data penelitian ini diperoleh dari dua sumber utama, yaitu BMKG dan Dinas Pertanian/BPS, dengan rentang waktu 2009 hingga 2024. Variabel cuaca yang dikumpulkan meliputi suhu rata-rata, curah hujan, kelembapan udara, serta lama penyinaran matahari. Variabel pertanian yang digunakan adalah luas panen, produktivitas, dan total produksi padi. Seluruh data diunduh secara resmi dan dicatat lengkap dengan metadata seperti tanggal unduh, satuan, dan cakupan wilayah. Sumber data ini dipilih karena kredibilitasnya dalam menyediakan data statistik yang akurat dan konsisten. Pemilihan variabel juga didasarkan pada relevansi agronomis yang telah terbukti dalam literatur terdahulu. Dengan landasan ini, pengumpulan data menjadi dasar untuk memastikan model *ANN* yang dibangun memiliki input berkualitas tinggi.

Proses pengumpulan data tidak hanya sebatas mengunduh, tetapi juga mencakup proses penyelarasan resolusi spasial dan temporal. Data cuaca yang biasanya tersedia dalam format harian atau bulanan diubah menjadi agregasi tahunan untuk diselaraskan dengan data hasil panen. Demikian pula, cakupan wilayah harus dipastikan sama, misalnya tingkat kabupaten atau kecamatan. Ketidaksesuaian format data antar-sumber dapat menimbulkan bias, sehingga tahap harmonisasi sangat penting dilakukan. Selain itu, catatan anomali iklim seperti El-Niño dan La-Niña diberi kode khusus untuk membantu model mengenali pengaruh kejadian ekstrem. Proses ini juga dilengkapi dengan pencatatan kelengkapan data untuk mengetahui tingkat missing values. Dengan demikian, tahap pengumpulan data bukan hanya aktivitas teknis, melainkan juga



bagian dari proses validasi awal. Tahapan ini memastikan bahwa dataset yang digunakan mencerminkan kondisi nyata dan siap untuk dianalisis lebih lanjut.

### 3.1.2 Pra-Pemrosesan Data (Preprocessing)

Tahap preprocessing bertujuan untuk mempersiapkan data mentah agar siap digunakan dalam pelatihan model (*ANN*). Salah satu proses utama dalam tahap ini adalah data cleaning yang mencakup penanganan nilai hilang, duplikasi, dan anomali. Missing values diatasi dengan metode imputasi mean, median, atau interpolasi linier pada data deret waktu. Data duplikat dihapus berdasarkan kombinasi variabel unik seperti tahun dan wilayah agar tidak menimbulkan bias. *Outlier* diidentifikasi menggunakan metode statistik seperti *Z-Score* dengan ambang  $|z| > 3$ . Nilai ekstrim yang masih relevan secara agronomis dipertahankan, sedangkan yang tidak valid dihapus. Seluruh langkah ini bertujuan untuk menjaga keutuhan dan representativitas dataset. Hasil akhirnya adalah data yang bersih, konsisten, dan siap masuk ke tahap transformasi. Sebelum digunakan dalam proses pelatihan model, dilakukan uji multikolinearitas antar variabel input untuk memastikan tidak terdapat hubungan linear yang terlalu tinggi. Uji ini dilakukan dengan menghitung nilai *Variance Inflation Factor* (VIF) menggunakan Persamaan 3.1 (Salmerón Gómez *et al.*, 2020):

$$VIF_i = \frac{1}{1-R_i^2} \quad (3.1)$$

Keterangan :

$VIF_i$  = Variance Inflation Factor untuk variabel independen ke- $i$

$R_i^2$  = Koefisien determinasi (R-squared) hasil regresi variabel  $i$

Di mana  $R_i^2$  adalah nilai koefisien determinasi hasil regresi variabel ke-i terhadap variabel independen lainnya. Jika nilai  $VIF_i > 10$ , maka variabel tersebut dianggap mengalami multikolinearitas tinggi. Hasil uji menunjukkan seluruh variabel memiliki nilai  $VIF < 10$  sehingga semua variabel layak digunakan sebagai input model *ANN*.

Selain pembersihan, preprocessing juga mencakup transformasi data agar sesuai dengan kebutuhan *ANN*. Normalisasi dengan metode *Z-Score* diterapkan untuk menyamakan skala antar variabel. Normalisasi ini dinyatakan dengan Persamaan (Permana and Salisah, 2022) :

$$X' = \frac{X_i - \mu}{\sigma} \quad (3.2)$$

Keterangan :

$X_i$  : nilai asli data ke-i

$\mu$  : rata-rata variabel

$\sigma$  : standar deviasi

Metode *Z-Score* dipilih karena mampu mempertahankan bentuk distribusi data asli dan membuat setiap variabel memiliki rata-rata nol dan simpangan baku satu. Proses ini penting karena *ANN* sensitif terhadap perbedaan skala antar fitur. Teknik ini memastikan bahwa variabel dengan rentang besar, seperti curah hujan, tidak mendominasi variabel lain seperti suhu. Variabel kategorikal, misalnya kecamatan, dikodekan menggunakan metode *One-Hot Encoding* agar dapat diproses oleh algoritma. Selain itu, penyelarasan rentang waktu dilakukan dengan mengagregasi data bulanan ke format tahunan. Transformasi ini juga mencakup penyamaan satuan pengukuran seperti suhu dalam °C dan curah hujan dalam mm. Semua langkah ini dirancang untuk meningkatkan stabilitas dan efisiensi pelatihan model. Dengan

preprocessing yang tepat, data menjadi lebih representatif dan dapat mendukung kinerja prediktif *ANN*.

### 3.1.3 Pemodelan (*Artificial Neural Network* – MLP)

Model utama yang digunakan dalam penelitian ini adalah *Artificial Neural Network* (*ANN*) dengan arsitektur *Multilayer Perceptron* (MLP). *ANN* dipilih karena kemampuannya menangkap pola non-linear yang kompleks antar variabel cuaca dan tanah. Arsitektur model terdiri dari input layer, beberapa hidden layer, dan output layer. Input layer menerima data variabel seperti suhu, curah hujan, kelembapan, luas panen, dan hasil panen sebelumnya. *Hidden layer* berfungsi menangkap representasi abstrak dari data dengan fungsi aktivasi ReLU. *Output layer* menghasilkan prediksi hasil panen dalam bentuk nilai kontinu. Model ini dilatih menggunakan algoritma *backpropagation* untuk memperbarui bobot jaringan. Dengan rancangan ini, *ANN* mampu memodelkan kompleksitas hubungan faktor lingkungan dengan hasil panen.

Proses pelatihan model dilakukan dengan membagi data menjadi subset pelatihan dan pengujian. Dataset pelatihan digunakan untuk mengoptimalkan bobot jaringan, sedangkan dataset pengujian untuk menguji kemampuan generalisasi. Untuk mencegah *overfitting*, teknik regularisasi seperti *dropout* dan *early stopping* diterapkan. Jumlah *neuron* dan layer ditentukan melalui eksperimen awal yang diuji menggunakan data validasi. Pemilihan fungsi aktivasi linear pada output layer bertujuan agar model dapat menghasilkan prediksi tanpa batasan tertentu. Proses ini dilengkapi dengan baseline model sederhana seperti regresi linear untuk pembandingan. Dengan evaluasi berlapis,

model yang dipilih adalah yang memberikan performa terbaik pada data uji. Pendekatan ini memastikan *ANN* tidak hanya akurat, tetapi juga stabil.

#### 3.1.4 Optimasi Model (*Hyperparameter Tuning*)

Optimasi model dilakukan untuk menentukan kombinasi *hyperparameter* terbaik dalam *ANN*. *Hyperparameter* yang diuji mencakup jumlah *neuron* per layer, *learning rate*, *batch size*, jumlah epoch, dan fungsi aktivasi. Proses tuning menggunakan metode *Grid Search* yang menguji semua kombinasi nilai *hyperparameter*. Setiap kombinasi diuji dengan melatih model pada dataset pelatihan dan divalidasi pada dataset pengujian. Evaluasi dilakukan menggunakan metrik MAE, RMSE, dan  $R^2$ . Tujuannya adalah menemukan kombinasi yang menghasilkan akurasi terbaik dan risiko *overfitting* paling kecil. Proses ini memerlukan waktu komputasi yang tinggi, tetapi hasilnya signifikan bagi stabilitas model. Dengan tuning, performa *ANN* meningkat secara nyata dibandingkan model default.

Selain *Grid Search*, penelitian ini juga mempertimbangkan pendekatan Random Search untuk eksplorasi awal. *Random Search* memungkinkan ruang pencarian lebih luas tanpa harus mencoba semua kombinasi. Hasil dari Random Search dapat menjadi acuan untuk mempersempit ruang pencarian *Grid Search*. *Hyperparameter* yang paling berpengaruh biasanya adalah *learning rate* dan jumlah *neuron*. Penyesuaian kecil pada parameter ini dapat berdampak besar terhadap akurasi prediksi. Evaluasi ulang pada kombinasi terbaik dilakukan untuk memastikan konsistensi hasil. Proses tuning ini kemudian menghasilkan model final dengan kinerja optimal. Dengan demikian, *ANN* yang digunakan lebih stabil, efisien, dan akurat.

### 3.1.5 Evaluasi Model

Evaluasi model bertujuan menilai sejauh mana *ANN* mampu memprediksi hasil panen dengan akurat. Tiga metrik utama digunakan, yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-Squared* ( $R^2$ ). MAE digunakan karena mudah diinterpretasikan dalam satuan ton hasil panen. RMSE dipilih karena memberikan penalti lebih besar pada kesalahan ekstrem.  $R^2$  digunakan untuk menilai sejauh mana model mampu menjelaskan variabilitas data. Selain metrik kuantitatif, visualisasi *Actual vs Predicted* juga dilakukan untuk memvalidasi kesesuaian pola. Evaluasi dilakukan pada data uji yang tidak pernah dilihat model selama pelatihan. Hal ini memastikan bahwa hasil evaluasi mencerminkan kemampuan generalisasi model.

Selain evaluasi kuantitatif, analisis residu juga dilakukan untuk mendeteksi bias sistematis. Residual plot membantu mengidentifikasi kesalahan prediksi yang cenderung konsisten pada periode tertentu. Analisis spasial juga dipertimbangkan dengan membandingkan performa model pada tiap kecamatan. Model yang baik harus stabil di berbagai wilayah, bukan hanya di wilayah dominan.

## 3.2 Desain Penelitian

Desain penelitian ini bertujuan untuk memprediksi hasil panen padi dengan menggunakan model (*ANN*) berdasarkan faktor-faktor lingkungan, seperti cuaca dan kondisi tanah. Penelitian ini dilakukan dalam beberapa tahapan mulai dari pengumpulan data hingga evaluasi hasil prediksi yang dihasilkan oleh model. Berikut adalah langkah-langkah utama dalam desain penelitian ini:

### 3.2.1 Pengumpulan dan Pemahaman Data

Pengumpulan data merupakan tahap awal yang sangat penting dalam proses penelitian ini karena kualitas dan relevansi data yang dikumpulkan akan memengaruhi hasil prediksi yang dihasilkan oleh model (*ANN*). Dalam penelitian ini, data yang digunakan berasal dari dua sumber utama, yaitu BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) dan dinas pertanian, dengan cakupan waktu antara tahun 2020 hingga 2024 dan wilayah fokus di Malang. Adapun jenis data yang dikumpulkan dibagi menjadi dua kategori utama, yaitu data cuaca dan data pertanian.

#### 1. Data Cuaca

Data ini diperoleh dari BMKG dan mencakup variabel-variabel lingkungan yang dapat memengaruhi hasil panen. Data cuaca dikumpulkan secara periodik (bulanan atau tahunan) dan terdiri dari:

- a) Suhu rata-rata udara ( $^{\circ}\text{C}$ )
- b) Kelembapan udara (%)
- c) Curah hujan (mm)

Data ini direkam selama periode waktu tertentu untuk mengetahui pola dan tren iklim yang berdampak pada pertanian.

#### 2. Data Pertanian

Data ini diperoleh dari dinas pertanian dan mencakup informasi spesifik mengenai kegiatan pertanian, terutama terkait produksi padi. Variabel-variabel yang dikumpulkan meliputi:

- a) Luas panen (hektar)
- b) Jumlah produksi padi (ton)
- c) Data pendukung lainnya terkait distribusi wilayah dan waktu panen.

Berdasarkan data yang telah dikumpulkan, penelitian ini mengelompokkan variabel menjadi dua kategori utama, yaitu variabel bebas dan variabel terikat. Variabel bebas merupakan variabel *input* yang diasumsikan memengaruhi hasil panen padi dan tidak dipengaruhi oleh variabel lain dalam konteks model prediksi. Variabel-variabel ini meliputi suhu udara rata-rata, kelembapan udara, curah hujan, dan luas panen. Seluruh variabel tersebut digunakan sebagai fitur *input* dalam model *ANN* karena secara ilmiah diketahui memiliki korelasi terhadap tingkat produksi padi. Sementara itu, variabel terikat dalam penelitian ini adalah jumlah produksi padi, yang menjadi *output* dari model prediksi. Nilai dari variabel ini diasumsikan dipengaruhi secara langsung oleh kombinasi nilai dari variabel bebas. Model *ANN* akan dilatih untuk memprediksi jumlah produksi padi berdasarkan data historis dari seluruh variabel bebas yang tersedia.

### 3.2.2 Data preprocessing

Data *preprocessing* merupakan tahapan krusial dalam penelitian ini yang bertujuan untuk mempersiapkan data mentah agar dapat digunakan secara optimal dalam proses pelatihan model *Artificial Neural Network (ANN)*. Tahapan ini terdiri dari empat bagian utama yang saling berkaitan, yaitu *data cleaning*, *data extraction*, *data transformation*, dan *feature selection*, sebagaimana tergambar dalam diagram alur penelitian.

a. *Data Cleaning*

Langkah pertama dalam proses *preprocessing* data adalah membersihkan data dari berbagai permasalahan kualitas yang dapat mengganggu kinerja model. Salah satu masalah umum yang sering ditemukan adalah adanya missing values atau nilai-nilai yang hilang. Untuk mengatasi hal ini, beberapa pendekatan digunakan sesuai dengan karakteristik data. Pada variabel numerik seperti suhu dan curah hujan, diterapkan metode imputasi mean, di mana nilai yang hilang digantikan dengan rata-rata dari nilai-nilai yang tersedia, yang secara matematis dirumuskan pada persamaan 3.3 (Lawson, 2023):

$$x_{baru} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.3)$$

di mana  $x_{baru}$  adalah nilai hasil imputasi (pengganti nilai hilang),  $x_i$  adalah nilai ke- $i$  dari variabel yang bersangkutan, dan  $n$  adalah jumlah total data yang tidak hilang (tersedia). Dengan cara ini, nilai yang hilang dapat diperkirakan secara statistik tanpa mengubah distribusi data secara drastis.

Selain itu, jika data bersifat deret waktu (*time-series*), maka metode interpolasi linier digunakan untuk memperkirakan nilai yang hilang berdasarkan tren data sebelumnya dan sesudahnya. Apabila jumlah nilai kosong sangat kecil dan dinilai tidak berdampak signifikan terhadap keseluruhan dataset, maka baris-baris yang mengandung nilai kosong tersebut dihapus secara selektif. Selain *missing values*, permasalahan lain yang ditangani dalam tahap ini adalah data duplikat dan anomali. Data duplikat dihapus agar tidak memengaruhi distribusi data, sementara data anomali dianalisis lebih lanjut untuk menentukan apakah perlu dikoreksi atau dihapus,



tergantung pada konteks dan signifikansinya terhadap variabel target. Proses deteksi duplikat dilakukan berdasarkan ID wilayah dan waktu. Data duplikat dihapus untuk menghindari bias. Anomali ekstrim (*outliers*) juga diidentifikasi menggunakan metode statistik seperti *Z-Score*, yaitu (Henderi 2021):

$$Z = \frac{X - \mu}{\sigma} \quad (3.4)$$

Keterangan :

$X$  adalah nilai asli dari fitur,

$\mu$  adalah rata-rata dari fitur tersebut,

$\sigma$  adalah standar deviasi dari fitur tersebut.

Jika nilai absolut *Z-Score* lebih dari 3 ( $|z| > 3$ ) (Henderi 2021), maka data tersebut dikaji ulang sebagai kemungkinan *outlier*.

## 1. Data Extraction

Data yang digunakan dalam proyek ini diperoleh dari dua sumber utama, yaitu BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) yang menyediakan informasi terkait kondisi cuaca seperti curah hujan, suhu udara, dan kelembapan, serta dari dinas pertanian yang menyediakan data hasil pertanian, termasuk luas panen dan jumlah produksi. Dalam tahap ekstraksi ini, perhatian khusus diberikan untuk menyamakan cakupan wilayah (misalnya kabupaten atau provinsi) dan periode waktu (misalnya bulanan atau tahunan), agar data yang digabungkan dari kedua sumber tersebut dapat digunakan secara konsisten dan relevan dalam proses analisis selanjutnya. Adapun proses yang penting untuk dilakukan yakni menyamakan format dan granularity waktu (tahunan) dan lokasi (misalnya per kabupaten atau provinsi). Kemudian dilakukan

penggabungan dengan melalui *inner join* berdasarkan tahun dan wilayah, memastikan hanya data yang lengkap dari kedua sumber yang digunakan.

## 2. Data Transformation

Setelah proses ekstraksi, data yang diperoleh kemudian melalui tahap transformasi untuk mengubah formatnya menjadi seragam dan siap digunakan dalam model. Misalnya, satuan pengukuran yang berbeda—seperti curah hujan dalam milimeter dan suhu dalam derajat *Celsius*—disesuaikan agar memiliki keseragaman. Selain itu, transformasi juga mencakup proses normalisasi data numerik agar berada dalam rentang yang seragam, serta encoding untuk data kategorikal agar dapat diproses oleh algoritma pembelajaran mesin. Transformasi ini bertujuan untuk meningkatkan efisiensi model dan meminimalkan bias akibat skala atau format data yang tidak konsisten.

### 3.2.3 Feature Selection

Tahap *feature selection* dilakukan untuk meningkatkan efisiensi dan akurasi model dengan hanya mempertahankan fitur-fitur yang relevan terhadap hasil panen. Fitur-fitur yang digunakan dalam penelitian ini meliputi suhu udara rata-rata, kelembapan udara, curah hujan, dan luas panen, yang seluruhnya dikategorikan sebagai variabel bebas. Sementara itu, jumlah hasil panen menjadi satu-satunya variabel terikat atau target.

Pemilihan fitur variabel dalam penelitian ini didasarkan pada hubungan kausal yang umum digunakan dalam pendekatan prediktif berbasis supervised learning. Variabel suhu udara rata-rata, kelembapan udara, curah hujan, dan luas panen

dikategorikan sebagai variabel bebas (*independen*) karena secara teoritis dan empiris mereka berperan sebagai faktor penyebab yang memengaruhi hasil produksi padi. Variabel-variabel ini digunakan sebagai input atau fitur dalam model (*ANN*), karena memiliki pengaruh signifikan terhadap proses fisiologis tanaman, ketersediaan air, serta luas lahan yang dapat ditanami. Sebagai contoh, curah hujan berlebih dapat menyebabkan genangan yang menurunkan produktivitas, sementara suhu yang ekstrem dapat menghambat pertumbuhan tanaman.

Di sisi lain, jumlah hasil panen padi ditetapkan sebagai variabel terikat (*dependen*) karena merupakan keluaran (*output*) yang nilainya dipengaruhi oleh kombinasi dari keempat variabel bebas tersebut. Dalam konteks machine learning, variabel ini bertindak sebagai target yang ingin diprediksi oleh model berdasarkan pola historis data input. Dengan demikian, klasifikasi ini mendukung kerangka prediktif dalam penelitian, di mana *ANN* dilatih untuk memahami hubungan kompleks antara faktor-faktor cuaca dan agronomis dengan output berupa hasil panen padi. Rincian fitur dan target yang digunakan disajikan dalam Tabel 3.1

Tabel 3.1 Variabel Terikat dan Bebas Penelitian

Kategori	Nama Variabel	Satuan
Fitur (Variabel Bebas)	Suhu udara rata-rata	°C
	Kelembapan udara	%
	Curah hujan	mm
	Luas panen	hektar (ha)
Target (Variabel Terikat)	Jumlah hasil panen	ton

Pada Tabel 3.1 menunjukkan daftar variabel yang digunakan dalam penelitian ini, yang terdiri dari variabel bebas (fitur) dan variabel terikat (target). Variabel bebas mencakup empat parameter utama yang berpengaruh terhadap hasil panen, yaitu suhu

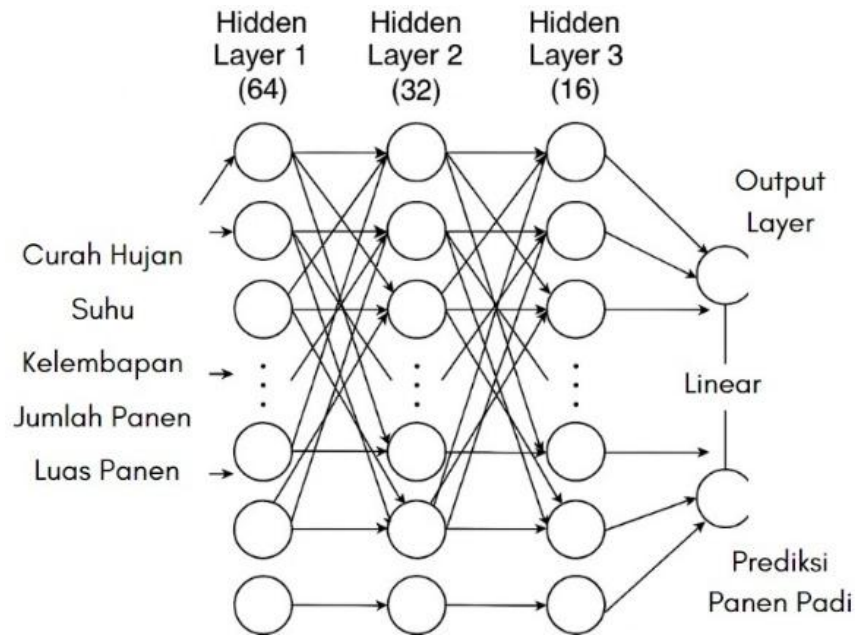
udara rata-rata (dalam satuan °C), kelembapan udara (dalam persen), curah hujan (dalam milimeter), dan luas panen (dalam hektar). Sementara itu, variabel terikat dalam penelitian ini adalah jumlah hasil panen yang diukur dalam satuan ton. Seluruh variabel ini digunakan sebagai *input* dan *output* dalam proses pelatihan model (*ANN*) untuk memprediksi hasil panen berdasarkan kondisi cuaca dan luas lahan yang tersedia.

Setelah seluruh proses *preprocessing* selesai, dataset akhir terdiri dari data numerik yang telah dibersihkan dari *noise* dan nilai hilang, baik melalui imputasi mean, interpolasi linier, maupun penghapusan baris yang tidak signifikan. Data juga telah dinormalisasi menggunakan metode *Z-Score*, yang bertujuan menyetarakan skala antar fitur agar model dapat memproses *input* dengan optimal.

Dataset akhir hanya mengandung fitur-fitur utama yang relevan dan siap untuk dibagi ke dalam data pelatihan dan pengujian. Setelah seluruh proses di atas, dataset akhir terdiri atas data numerik yang bersih dari *noise* dan nilai hilang, telah dinormalisasi (*Z-Score*), mengandung fitur-fitur utama yang relevan, dan siap untuk dibagi ke dalam data pelatihan dan pengujian dataset ini kemudian diproses ke tahap *Split Data* dan *Modelling* sebagaimana terlihat dalam diagram alur penelitian.

### 3.2.4 Modelling menggunakan Algoritma *ANN*

Pada tahap ini, dibangun model prediksi hasil panen padi menggunakan algoritma (*ANN*) dengan pendekatan *Multilayer Perceptron* (MLP). *ANN* digunakan karena kemampuannya dalam memetakan hubungan non-linear yang kompleks antara *input* (seperti faktor cuaca dan lahan) dengan *output* berupa hasil panen. Arsitektur *ANN* yang digunakan dapat dilihat pada Gambar 3.2 berikut:



Gambar 3.2 Arsitektur ANN pada Prediksi Panen Padi

Gambar 3.2 Secara matematis, proses propagasi maju (*forward propagation*) pada (ANN) dapat dinyatakan dengan Persamaan 3.6 berikut (Irvanizam, Horatius, and Sofyan, 2023):

$$a_j^{(l)} = f(\sum_{i=1}^n w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)}) \quad (3.6)$$

$a_j^{(l)}$  = aktivasi *neuron* ke- $j$  pada lapisan ke- $l$

$w_{ij}^{(l)}$  = bobot dari *neuron* ke- $i$  pada lapisan sebelumnya ke *neuron* ke- $j$

$a_i^{(l-1)}$  = output *neuron* pada lapisan sebelumnya

$b_j^{(l)}$  = bias dari *neuron* ke- $j$  pada lapisan ke- $l$

$f$  = fungsi aktivasi (*activation function*)

$l$  = indeks lapisan dalam jaringan saraf.

$i$  = indeks *neuron* pada lapisan sebelumnya

$j$  = indeks *neuron* pada lapisan saat ini

Persamaan ini menggambarkan bagaimana setiap *neuron* menerima sejumlah input yang dikalikan dengan bobot, ditambahkan dengan bias, lalu hasilnya diproses menggunakan fungsi aktivasi.

Pada penelitian ini, fungsi aktivasi yang digunakan di hidden layer adalah Rectified Linear Unit (ReLU), yang dinyatakan pada Persamaan 3.7 (Urenda and Kreinovich, 2022):

$$f(x) = \max(0, x) \quad (3.7)$$

Fungsi ReLU dipilih karena mampu memodelkan hubungan non-linear antara faktor cuaca dan kondisi tanah terhadap hasil panen padi. ReLU juga mempercepat proses pelatihan dan membantu menghindari masalah *vanishing gradient*.

Fungsi aktivasi ReLU digunakan untuk memperkenalkan *non-linearitas* ke dalam jaringan saraf sehingga model dapat mempelajari hubungan yang kompleks antara input dan output. ReLU bekerja dengan prinsip sederhana, yaitu memberikan output sebesar nol untuk semua nilai negatif dan mempertahankan nilai positif apa adanya. Secara matematis, fungsi ini mudah dihitung, yang menjadikannya lebih efisien secara komputasi dibanding fungsi aktivasi lain seperti sigmoid dan tanh.

Dibandingkan fungsi sigmoid, yang memiliki rentang output antara 0 dan 1, ReLU tidak menyebabkan vanishing gradient problem secara signifikan karena gradien tetap konstan (1) untuk nilai positif. Dengan demikian, proses pelatihan menjadi lebih cepat dan stabil, terutama pada model dengan banyak lapisan tersembunyi. Selain itu, ReLU juga membantu model untuk melakukan sparse activation, yaitu hanya sebagian *neuron* yang aktif pada satu waktu, sehingga mengurangi beban komputasi dan risiko *overfitting*.

Dalam konteks penelitian ini, penggunaan ReLU penting karena hubungan antara faktor cuaca (seperti suhu, curah hujan, kelembapan) dan kondisi tanah (seperti pH dan kelembapan tanah) terhadap hasil panen padi bersifat non-linear. Fungsi ReLU memungkinkan model *ANN* menangkap pola-pola non-linear tersebut secara efektif, meningkatkan akurasi prediksi hasil panen dibanding jika menggunakan fungsi aktivasi linier atau sigmoid.

Sementara itu, pada lapisan output digunakan fungsi aktivasi linear, sebagaimana ditunjukkan pada Persamaan 3.8 (Altunöz, 2024):

$$\hat{y} = \sum_{i=1}^n w_i a_i + b \quad (3.8)$$

Keterangan :

$\hat{y}$  : Merupakan output atau prediksi akhir dari model  
 $n$  : Jumlah *neuron* pada lapisan sebelumnya (biasanya hidden layer terakhir).  
 $w_i$  : Bobot (weight) dari *neuron* ke-iii yang menghubungkan *neuron* pada hidden layer dengan *neuron* output.  
 $a_i$  : Nilai aktivasi dari *neuron* ke-iii pada lapisan sebelumnya (hidden layer).  
 $b$  : Bias, yaitu nilai tambahan yang memungkinkan model untuk melakukan pergeseran terhadap garis regresi agar tidak selalu melalui titik nol.

Persamaan tersebut menggambarkan bahwa output model ( $\hat{y}$ ) dihitung sebagai jumlah dari seluruh aktivasi *neuron* pada lapisan sebelumnya yang telah dikalikan dengan bobot masing-masing. Setiap *neuron* pada hidden layer terakhir menghasilkan nilai aktivasi  $a_i$ , dan tiap nilai tersebut memiliki bobot  $w_i$  yang menunjukkan seberapa besar pengaruhnya terhadap output. Kemudian, sebuah nilai tambahan berupa bias ( $b$ ) disertakan agar model memiliki fleksibilitas lebih dalam menyesuaikan garis prediksi

bias ini membuat model tidak harus selalu melewati titik nol. Dengan demikian, output model merupakan kombinasi linear dari bobot, aktivasi, dan bias.

Secara sederhana, proses pada lapisan output dapat diibaratkan seperti menjumlahkan tiga komponen utama: nilai aktivasi dari *neuron-neuron* sebelumnya, bobot yang menunjukkan seberapa besar pengaruh masing-masing aktivasi, dan sebuah nilai bias yang memberikan penyesuaian tambahan. Jika dianalogikan dapat seperti penjumlahan  $a + b + x$ , maka (a) mewakili kontribusi aktivasi *neuron* pertama setelah dikalikan bobotnya, (b) mewakili kontribusi *neuron* kedua, dan (x) adalah nilai bias yang ditambahkan untuk menggeser hasil akhir agar model lebih fleksibel. Ketiga elemen tersebut digabungkan sehingga menghasilkan satu nilai output yang menjadi prediksi akhir dari model. Dengan kata lain, output *ANN* terbentuk dari penjumlahan linear berbagai komponen yang masing-masing memiliki peran dalam membentuk nilai prediksi.

Fungsi linear digunakan karena hasil prediksi berupa nilai kontinu (jumlah hasil panen dalam ton). Dengan kombinasi ReLU pada hidden layer (non-linear) dan fungsi linear pada output layer, model *ANN* dapat mengenali pola non-linear yang kompleks dari data, namun tetap menghasilkan output kuantitatif yang realistis dan dapat diinterpretasikan. Setelah proses *forward propagation* menghasilkan nilai prediksi  $\hat{y}$ , langkah berikutnya adalah menghitung error atau selisih antara nilai aktual (y) dan nilai prediksi ( $\hat{y}$ ) untuk memperbarui bobot jaringan. Proses pembaruan ini disebut sebagai *backpropagation*.



Pada tahap pelatihan, model *ANN* melakukan iterasi berulang untuk meminimalkan error melalui tiga langkah utama, yaitu:

- (1) *Forward propagation*,
- (2) *Backward propagation*, dan
- (3) *Weight update*.

### 1. **Forward Propagation**

Pada tahap ini, data masukan diproses melalui jaringan untuk menghasilkan output prediksi. Output akhir dari jaringan adalah nilai prediksi hasil panen ( $\hat{y}$ ) yang dihasilkan oleh fungsi linear pada output layer.

### 2. **Backward Propagation (Perhitungan Gradien)**

Setelah nilai prediksi diperoleh, model menghitung error menggunakan fungsi *loss*. Dalam penelitian ini digunakan fungsi *Mean Squared Error (MSE)* sebagaimana Persamaan 3.9 (Karunasingha, 2022b):

$$L = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.9)$$

Keterangan

$L$  = Nilai Loss Function

$n$  = Jumlah total data atau banyaknya pengamatan (sampel)

$y_i$  = Nilai aktual (nilai sebenarnya) dari data ke- $i$ .

$\hat{y}_i$  = Nilai prediksi (nilai hasil model) dari data ke- $i$ .

Fungsi *loss* ini mengukur seberapa jauh hasil prediksi berbeda dari nilai sebenarnya. Semakin kecil nilai *loss*, semakin baik performa model.

Nilai gradien dari *loss function* terhadap setiap bobot ( $w_{ij\_}\{ij\}w_{ij}$ ) dihitung

menggunakan turunan parsial, kemudian digunakan untuk memperbarui bobot melalui algoritma Gradient Descent.

### 3. Pembaruan Bobot (Weight Update)

Pembaruan bobot dilakukan dengan menyesuaikan setiap bobot berdasarkan arah negatif gradien, sebagaimana dinyatakan pada Persamaan 3.10 berikut (Larwuy, 2024):

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \eta \frac{\partial L}{\partial w_{ij}^{(l)}} \quad (3.10)$$

$\eta$ : *learning rate* (tingkat pembelajaran)

$\frac{\partial L}{\partial w_{ij}^{(l)}}$ : turunan fungsi *loss* terhadap bobot ke- $i,j$  pada layer ke- $l$

Persamaan tersebut menunjukkan bahwa setiap bobot dikoreksi secara bertahap agar nilai prediksi semakin mendekati nilai aktual. Nilai *learning rate* menentukan seberapa besar langkah perubahan bobot pada setiap iterasi. Nilai yang terlalu besar dapat menyebabkan model tidak stabil, sedangkan nilai yang terlalu kecil membuat proses pelatihan lambat.

Proses *forward* dan *backward propagation* diulang untuk setiap epoch hingga nilai *loss function* konvergen (tidak banyak berubah). Dengan cara ini, model *ANN* mampu mempelajari pola hubungan kompleks antara variabel cuaca dan kondisi tanah terhadap hasil panen padi.

Kemudian jaringan di atas menggambarkan alur data dari *input* hingga prediksi. Setiap lapisan dihubungkan penuh ke lapisan berikutnya (*fully connected*),

menunjukkan bahwa setiap *input* dapat memengaruhi seluruh *neuron* di *layer* berikutnya. Arsitektur ini dirancang untuk menangkap hubungan kompleks yang tidak dapat ditangkap oleh model linear biasa.

### 1. Arsitektur Jaringan

Model *Artificial Neural Network (ANN)* yang digunakan dalam penelitian ini dirancang dengan arsitektur berlapis (*layered architecture*) yang terdiri dari beberapa komponen utama, yakni *input layer*, *Hidden Layers*, dan *output layer*. Desain arsitektur ini bertujuan untuk menangkap hubungan non-linear antara fitur-fitur *input* dengan target *output*, serta meningkatkan kemampuan generalisasi model dalam memprediksi hasil panen berdasarkan data historis. *Input Layer*: Terdiri dari lima *neuron*, masing-masing merepresentasikan variabel *input* yaitu *curah hujan*, *suhu*, *kelembapan*, *jumlah panen sebelumnya*, dan *luas panen*.

#### a) *Input layer*

Lapisan *input* terdiri dari lima *neuron*, yang masing-masing mewakili satu variabel *input* utama yang digunakan dalam proses prediksi. Kelima variabel tersebut meliputi:

- 1) Curah hujan (mm)
- 2) Suhu udara rata-rata (°C)
- 3) Kelembapan udara (%)
- 4) Jumlah hasil panen pada periode sebelumnya (ton) – digunakan sebagai informasi historis untuk membantu model memahami tren waktu.

5) Luas panen (hektar)

Setiap *neuron* pada *input layer* menerima satu fitur dari dataset yang telah dinormalisasi, kemudian mengirimkan sinyal tersebut ke lapisan tersembunyi untuk diproses lebih lanjut.

b) *Hidden Layer*

Jaringan ini menggunakan tiga lapisan tersembunyi (*Hidden Layers*) yang masing-masing terdiri dari sejumlah *neuron* yang berbeda. Jumlah *neuron* pada setiap lapisan dipilih berdasarkan hasil eksperimen awal dan *tuning* untuk mencapai keseimbangan antara kompleksitas model dan performa prediksi. Adapun untuk konfigurasi *Hidden Layer* yang dipakai yakni seperti :

1) *Hidden Layer 1:*

Terdiri dari 64 *neuron*. Lapisan ini berfungsi untuk menangkap pola kompleks awal dari kombinasi fitur *input*. Aktivasi yang digunakan adalah fungsi ReLU (*Rectified Linear Unit*), yang memungkinkan model menangani hubungan non-linier dengan lebih baik dan menghindari masalah *vanishing gradient*.

2) *Hidden Layer 2:*

Terdiri dari 32 *neuron*. Lapisan ini berperan menyaring dan memperdalam pemahaman terhadap fitur-fitur hasil ekstraksi dari *layer* sebelumnya.

3) *Hidden Layer 3:*

Terdiri dari 16 *neuron*. Lapisan ini menyusun representasi akhir dari fitur sebelum diteruskan ke *output layer*. Pengurangan jumlah *neuron* secara bertahap bertujuan untuk menghindari *overfitting* dan menyederhanakan struktur model.

Setiap *neuron* dalam *Hidden Layer* menghitung aktivasi berdasarkan bobot *input* dan bias menggunakan fungsi aktivasi, biasanya ReLU (Do and Prodan 2024):

$$a_j^{(l)} = \text{ReLU}(\sum_{i=1}^n w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)}) \quad (3.11)$$

Keterangan :

$a_j^{(l)}$ : aktivasi *neuron* ke- $j$  pada *layer* ke- $l$

$w_{ij}^{(l)}$  bobot dari *neuron* ke- $i$  di *layer* sebelumnya ke *neuron* ke- $j$  di *layer* ke- $l$

$b_j^{(l)}$ : bias *neuron* ke- $j$

$\text{ReLU}(x) = \max(0, x)$

$l$  = indeks lapisan dalam jaringan saraf.

$i$  = indeks *neuron* pada lapisan sebelumnya

$j$  = indeks *neuron* pada lapisan saat ini

Setiap *neuron* pada *Hidden Layers* akan melakukan kalkulasi bobot terhubung dengan *neuron* dari lapisan sebelumnya, ditambahkan dengan bias, kemudian dilewatkan ke fungsi aktivasi. Fungsi aktivasi ReLU digunakan di semua *Hidden Layer* untuk memastikan aktivasi non-linier tetap efisien dan stabil.

#### c) *Output layer*

*Output layer* terdiri dari satu *neuron* yang merepresentasikan prediksi jumlah hasil panen (ton). Karena *output* yang diharapkan bersifat kontinu, digunakan fungsi

aktivasi linear pada *output layer* agar nilai prediksi tidak dibatasi dalam kisaran tertentu. Fungsi aktivasi yang digunakan pada *layer* ini adalah linear (Altunöz 2024):

$$\hat{y} = \sum_{i=1}^n w_i a_i + b \quad (3.12)$$

Keterangan :

$\hat{y}$  : Merupakan output atau prediksi akhir dari model

$n$  : Jumlah *neuron* pada lapisan sebelumnya (biasanya hidden layer terakhir).

$w_i$  : Bobot (weight) dari *neuron* ke-iii yang menghubungkan *neuron* pada hidden layer dengan *neuron* output.

$a_i$  : Nilai aktivasi dari *neuron* ke-iii pada lapisan sebelumnya (hidden layer).

$b$  : Bias, yaitu nilai tambahan yang memungkinkan model untuk melakukan pergeseran terhadap garis regresi agar tidak selalu melalui titik nol.

Rumus ini menggambarkan bahwa output akhir dari jaringan saraf adalah kombinasi linear dari semua aktivasi yang diterima dari lapisan sebelumnya, dikalikan dengan bobot masing-masing, lalu ditambahkan bias. Karena fungsi aktivasi linear tidak membatasi nilai output dalam rentang tertentu, maka cocok digunakan untuk kasus regresi seperti prediksi hasil panen yang nilainya berupa angka kontinu dan tidak terbatas (misalnya dari 0 hingga puluhan ribu ton). Penggunaan fungsi aktivasi linear ini bertujuan untuk menjaga keluaran model tetap dalam skala asli dari target prediksi, sehingga interpretasi hasilnya relevan dengan data sebenarnya

## 2. Simulasi Pemodelan ANN

Dalam simulasi perhitungan ini, digunakan sebuah jaringan saraf sederhana yang terdiri dari tiga *neuron* pada lapisan *input* untuk mewakili masing-masing faktor cuaca, dua *neuron* pada lapisan tersembunyi (*Hidden Layer*), dan satu *neuron* pada lapisan *output* untuk menghasilkan prediksi. Setiap *neuron* dalam *Hidden Layer* menerima kombinasi linier dari *input* yang dikalikan bobot tertentu dan ditambahkan dengan bias, kemudian hasilnya diproses melalui fungsi aktivasi sigmoid. Misalnya, untuk *neuron*

pertama pada *Hidden Layer*, kombinasi *input* 0.7 (curah hujan), 0.6 (kelembapan), dan 0.8 (suhu) dengan bobot masing-masing 0.2, 0.4, dan 0.1 serta bias 0.1.

Simulasi Perhitungan Aktivasi pada *Hidden Layer* 1

$$\begin{aligned} Z_{H1} &= (0.7 \times 0.2) + (0.6 \times 0.4) + (0.8 \times 0.1) + 0.1 \\ &= 0.14 + 0.24 + 0.08 + 0.1 = 0.56 \end{aligned} \quad (3.13)$$

$$a_{H1} = \text{sigmoid}(0.56) = \frac{1}{1 + e^{-0.56}} \approx 0.636$$

Pada Persamaan 3.13 merupakan contoh perhitungan yang dilakukan pada *Hidden Layer* pertama kemudian akan dilakukan perhitungan *Hidden Layer* kedua serta *output layer* yang menghasilkan nilai *output* berupa jumlah produksi padi dalam bentuk Ton.

$$\begin{aligned} Z_{H2} &= (0.7 \times 0.5) + (0.6 \times -0.3) + (0.8 \times 0.2) - 0.2 \\ &= 0.35 - 0.18 + 0.16 - 0.2 = 0.13 \end{aligned} \quad (3.14)$$

$$a_{H2} = \text{sigmoid}(0.13) = \frac{1}{1 + e^{-0.13}} \approx 0.532$$

Kemudian dilakukan perhitungan *output layer* berada pada persamaan 3.15 dibawah :

$$\begin{aligned}
 Z_{out} &= (0.636 \times 0.6) + (0.532 \times 0.3) + 0.05 \\
 &= 0.3816 + 0.1596 + 0.05 = 0.5912
 \end{aligned} \tag{3.15}$$

$$a_{out} = \text{sigmoid}(0.5912) = \frac{1}{1 + e^{-0.5912}} \approx 0.643$$

Nilai *output* dari *ANN* adalah 0.643 dalam skala 0–1.

Jika skala hasil panen adalah 0–10 ton/ha, maka:

$$\text{Hasil Panen} = 0.643 \times 10 = 6.43 \text{ ton/ha} \tag{3.16}$$

### 3. Langkah-langkah Pelatihan Model *ANN*

Proses pelatihan model *Artificial Neural Network (ANN)* dilakukan secara iteratif menggunakan algoritma gradient descent hingga diperoleh bobot optimal yang menghasilkan nilai error minimum. Tahapan proses pelatihan model dapat dijelaskan sebagai berikut:

#### a) Inisialisasi Bobot dan Bias

Semua bobot (*w*) dan bias (*b*) pada jaringan diinisialisasi dengan nilai acak yang sangat kecil. Langkah ini bertujuan agar semua *neuron* mulai belajar tanpa bias terhadap arah tertentu.

#### b) Forward Propagation

Data input diproses dari layer input menuju output. Setiap *neuron* pada hidden layer menghitung aktivasi berdasarkan kombinasi linier dari input dan bobot sesuai dengan Persamaan 3.11, lalu diaplikasikan fungsi aktivasi ReLU (Persamaan 3.12). Hasil akhir



pada output layer dihitung menggunakan fungsi linear untuk menghasilkan nilai prediksi.

c) Perhitungan Error (Loss Function)

Nilai error dihitung dengan membandingkan hasil prediksi dan nilai aktual menggunakan fungsi *Mean Squared Error (MSE)*. Semakin kecil nilai *loss*, semakin baik model dalam melakukan prediksi.

d) Backward Propagation

Error yang dihasilkan dikembalikan ke lapisan-lapisan sebelumnya untuk menghitung gradien dari setiap bobot menggunakan turunan parsial dari fungsi *loss*. Proses ini menentukan seberapa besar kontribusi setiap bobot terhadap error total.

e) Pembaruan Bobot (Weight Update)

Semua bobot dan bias diperbarui berdasarkan arah negatif gradien menggunakan algoritma *gradient descent*. Nilai *learning rate* ( $\eta$ ) mengontrol besarnya langkah pembaruan pada setiap iterasi.

f) Iterasi hingga Konvergen

Langkah 2 sampai 5 diulangi untuk setiap epoch hingga nilai *loss* konvergen atau jumlah iterasi maksimum tercapai. Konvergensi terjadi ketika perubahan nilai *loss* antar iterasi menjadi sangat kecil.

g) Pengujian Model

Setelah pelatihan selesai, model diuji menggunakan data uji untuk mengukur kinerja prediksi berdasarkan metrik MAE dan RMSE. Evaluasi ini memastikan bahwa model tidak hanya akurat terhadap data pelatihan, tetapi juga memiliki kemampuan generalisasi yang baik terhadap data baru.

### 3.2.5 Optimasi model (*Hyperparameter tuning*)

Tahap optimasi model bertujuan untuk memperoleh konfigurasi *hyperparameter* terbaik guna memaksimalkan performa prediksi dari model *Artificial Neural Network* (*ANN*). *Hyperparameter* merupakan parameter yang nilainya ditentukan sebelum proses pelatihan dimulai dan tidak diubah selama pelatihan. Pemilihan nilai yang tepat untuk *hyperparameter* sangat krusial karena secara signifikan dapat memengaruhi kualitas hasil prediksi model. Beberapa *hyperparameter* utama yang berperan penting dalam performa model *Artificial Neural Network* (*ANN*) antara lain mencakup: jumlah lapisan tersembunyi (*Hidden Layers*) yang menentukan kedalaman jaringan dan kompleksitas representasi; jumlah *neuron* per lapisan yang memengaruhi kapasitas model dalam menangkap pola dari data; *Learning rate*, yaitu besarnya langkah pembaruan bobot pada setiap iterasi selama proses pelatihan; jumlah *Epoch* yang menunjukkan berapa kali seluruh dataset digunakan untuk melatih model secara penuh; serta ukuran batch (*Batch size*) yang menentukan jumlah sampel yang digunakan dalam satu kali proses pembaruan bobot. Selain itu, pemilihan fungsi aktivasi (*activation function*) seperti ReLU, sigmoid, atau tanh memengaruhi kemampuan model untuk mempelajari hubungan non-linear, dan pemilihan *optimizer* seperti *Stochastic Gradient Descent* (*SGD*) atau Adam sangat memengaruhi efisiensi dan stabilitas proses pelatihan. Kombinasi nilai *hyperparameter* ini sangat menentukan kinerja akhir dari model *ANN*.

Dalam penelitian ini, proses *hyperparameter tuning* dilakukan menggunakan metode *Grid Search*, yaitu teknik eksplorasi yang menguji semua kombinasi yang

mungkin dari nilai-nilai *hyperparameter* yang telah ditentukan. Tujuan dari *Grid Search* adalah untuk menemukan kombinasi optimal yang menghasilkan performa terbaik berdasarkan metrik evaluasi. Secara matematis, *Grid Search* dapat didefinisikan sebagai proses optimasi terhadap ruang parameter  $\Theta$ , di mana:

$$\Theta = \theta_1 \times \theta_2 \times \dots \times \theta_n \quad (3.17)$$

dengan  $\theta_i$  adalah himpunan nilai yang mungkin untuk *hyperparameter* ke-i. Adapun untuk Langkah-langkah *Grid Search* sebagai berikut :

1. Menentukan ruang pencarian hyperparameter:

Misalnya, kita memiliki beberapa hyperparameter yang ingin diuji dalam model *ANN*:

- Jumlah *Neuron*: [32, 64, 128]
- *Learning rate*: [0.001, 0.01]
- *Batch size*: [16, 32, 64]
- *Epochs*: [100, 200]

*Grid Search* akan menguji semua kombinasi dari nilai-nilai hyperparameter tersebut.

$$TotalKombinasi = (Neuron) \times (Learning Rate) \times (Batch Size) \times (Epoch) \quad (3.18)$$

Jika kita memiliki 3 pilihan jumlah *neuron*, 2 pilihan *learning rate*, 3 pilihan *batch size*, dan 2 pilihan *epochs*, maka jumlah total kombinasi adalah:

$$3 \times 2 \times 3 \times 2 = 36 \quad (3.19)$$

Melatih model untuk setiap kombinasi:

Untuk setiap kombinasi dari hyperparameter, model *ANN* akan dilatih pada dataset pelatihan. Misalnya:

- Kombinasi pertama: Jumlah *Neuron* = 32, *Learning rate* = 0.001, *Batch size* = 16, *Epochs* = 100
- Kombinasi kedua: Jumlah *Neuron* = 32, *Learning rate* = 0.001, *Batch size* = 16, *Epochs* = 200
- Kombinasi ketiga: Jumlah *Neuron* = 32, *Learning rate* = 0.01, *Batch size* = 16, *Epochs* = 100
- Dan seterusnya hingga seluruh kombinasi diuji.

Setiap model dilatih menggunakan data pelatihan dan hasil evaluasi akan dihitung untuk setiap kombinasi. Setelah model dilatih, setiap kombinasi hyperparameter dievaluasi menggunakan metrik evaluasi seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), atau *R-Squared* ( $R^2$ ).

Tujuan *Grid Search* adalah menemukan (Jiang and Xu 2022) :

$$\theta^* = \arg \min \mathcal{L}(f(x; \theta), y) \quad (3.20)$$

Keterangan :

$f(x; \theta)$  adalah model *ANN* dengan parameter  $\theta$

$\mathcal{L}$  adalah fungsi loss atau metrik evaluasi, misalnya MSE, RMSE, MAE

X adalah data *input*, dan y adalah label aktual

Kemudian untuk langkah-langkah proses optimasi pada metode *Grid Search* yakni :

1. Menentukan ruang pencarian (*parameter space*)

*Parameter space* adalah area di mana kita mendefinisikan rentang nilai untuk setiap hyperparameter yang ingin dioptimasi. Tujuan dari langkah ini adalah untuk mencari kombinasi terbaik dari parameter tersebut, yang nantinya akan menghasilkan model dengan performa terbaik. Pada umumnya, hyperparameter dalam model neural network seperti *Artificial Neural Network (ANN)* meliputi:

a. **Jumlah Neuron:** Mengacu pada jumlah *neuron* yang akan digunakan di lapisan tersembunyi. Jumlah *neuron* yang tepat dapat mempengaruhi kemampuan model untuk menangkap pola data yang lebih kompleks. Pilihan yang sering dicoba meliputi:

[32, 64, 128, 256]

b. **Learning rate:** Menentukan seberapa besar langkah pembaruan bobot pada setiap iterasi pelatihan. *Learning rate* yang terlalu tinggi dapat menyebabkan model melewati titik minimum optimal, sementara yang terlalu rendah dapat memperlambat proses konvergensi. Pilihan yang sering dicoba termasuk:

[0.001, 0.01, 0.1]

c. **Batch size:** Menentukan jumlah sampel yang diproses dalam setiap iterasi pembaruan bobot. *Batch size* kecil sering lebih efektif untuk generalisasi, namun membutuhkan lebih banyak waktu. Pilihan yang sering dicoba adalah:

[16, 32, 64, 128]

d. **Epoch:** Menentukan jumlah iterasi di mana seluruh dataset digunakan untuk melatih model. Terlalu sedikit epoch bisa menyebabkan model *underfitting*, sementara terlalu banyak bisa menyebabkan *overfitting*. Pilihan umum termasuk:

[100, 200, 500]

e. **Optimizer:** Algoritma yang digunakan untuk mengoptimalkan model, seperti Adam, SGD, dan RMSprop. Penggunaan optimizer yang tepat dapat mempengaruhi kecepatan konvergensi dan kualitas hasil akhir. Pilihan umum:

['adam', 'sgd']

Untuk menemukan kombinasi hyperparameter yang optimal, kita dapat menggunakan metode *Grid Search*, yang akan mencoba semua kombinasi parameter yang mungkin dalam ruang pencarian yang telah ditentukan. Tabel 3.2 *Pseudocode Gridsearch* dalam menentukan parameter adalah *pseudocode* yang menggambarkan langkah-langkah dari proses *Grid Search* dalam optimasi *hyperparameter* model *ANN*.

**START**

**# Langkah 1: Tentukan Parameter yang Akan Dioptimasi**

# Definisikan ruang pencarian hyperparameter untuk model *ANN*

parameter\_space = {

  'jumlah\_neuron': [32, 64, 128], # Pilihan jumlah *neuron*

  'learning\_rate': [0.001, 0.01, 0.1], # Pilihan *learning rate*

  'batch\_size': [16, 32, 64], # Pilihan ukuran batch

  'epoch': [100, 200, 500], # Pilihan jumlah epoch

  'optimizer': ['adam', 'sgd'] # Pilihan optimizer

```

    }
# Langkah 2: Generate Semua Kombinasi Parameter yang Mungkin
    FUNCTION generate_combinations(parameter_space)
        combinations = [] # Inisialisasi list untuk menyimpan kombinasi parameter
        # Menghasilkan produk kartesian untuk semua parameter
        FOR each combination IN cartesian_product(parameter_space)
            combinations.append(combination)
        END FOR
        RETURN combinations
    END FUNCTION

# Langkah 3: Buat dan Latih Model dengan Kombinasi Parameter
    FUNCTION evaluate_model(X_train, y_train, combination)
        # Buat model ANN dengan parameter yang diberikan
        model = create_ANN_model(combination)
        # Latih model menggunakan data pelatihan
        model.fit(X_train, y_train)
        # Evaluasi model dan hitung skor kinerjanya
        score = evaluate_model_performance(model)
        RETURN score
    END FUNCTION

# Langkah 4: Menentukan Kombinasi Parameter Terbaik Berdasarkan Kinerja
    FUNCTION find_best_combination(X_train, y_train, parameter_space)
        best_score = -infinity # Inisialisasi skor terbaik
        best_combination = None # Inisialisasi kombinasi terbaik
        # Dapatkan semua kombinasi dari ruang pencarian
        combinations = generate_combinations(parameter_space)

        # Evaluasi setiap kombinasi parameter
        FOR each combination IN combinations
            score = evaluate_model(X_train, y_train, combination) # Evaluasi model untuk kombinasi ini
            # Jika skor lebih baik, simpan kombinasi ini sebagai yang terbaik
            IF score > best_score
                best_score = score
                best_combination = combination
            END IF
        END FOR

        RETURN best_combination, best_score
    END FUNCTION

# Langkah 5: Menjalankan Grid Search untuk Menemukan Parameter Terbaik
    # Eksekusi Grid Search dengan evaluasi semua kombinasi
    best_combination, best_score = find_best_combination(X_train, y_train, parameter_space)

    # Tampilkan hasil terbaik
    PRINT "Best Hyperparameters: ", best_combination
    PRINT "Best Performance Score: ", best_score

END

```

Pada Tabel 3.2 menggambarkan implementasi proses *Grid Search* untuk optimasi hyperparameter dalam model *Artificial Neural Network (ANN)*. Langkah pertama dalam proses ini adalah mendefinisikan ruang pencarian hyperparameter yang meliputi lima parameter utama yang mempengaruhi pelatihan model, yaitu jumlah *neuron*, *learning rate*, *batch size*, *epoch*, dan optimizer. Setiap parameter diberikan beberapa pilihan nilai yang akan diuji untuk mencari kombinasi terbaik yang memberikan kinerja optimal.

Setelah ruang pencarian ditentukan, langkah berikutnya adalah `generate_combinations`, yang menggunakan produk kartesian untuk menghasilkan semua kombinasi yang mungkin dari nilai-nilai hyperparameter yang telah ditentukan. Dengan demikian, semua konfigurasi yang mungkin akan diuji selama pencarian.

Selanjutnya, dalam fungsi `evaluate_model`, setiap kombinasi parameter yang dihasilkan akan digunakan untuk melatih model *ANN* menggunakan data pelatihan. Model tersebut kemudian dievaluasi menggunakan metrik kinerja tertentu, seperti *Mean Absolute Error (MAE)* atau *Root Mean Squared Error (RMSE)*, yang memberikan gambaran mengenai seberapa baik model bekerja. Skor yang diperoleh akan digunakan untuk membandingkan kinerja model dengan konfigurasi parameter lainnya.

Fungsi `find_best_combination` kemudian digunakan untuk mencari kombinasi parameter yang menghasilkan kinerja terbaik. Fungsi ini akan memeriksa setiap kombinasi yang dihasilkan, menguji model dengan kombinasi tersebut, dan memilih yang memberikan skor terbaik.



Akhirnya, setelah semua kombinasi diuji, *Grid Search* akan menghasilkan kombinasi hyperparameter terbaik dan skor kinerja terbaik, yang dapat digunakan untuk melatih model *ANN* yang lebih optimal. Hasil terbaik ini kemudian ditampilkan untuk memberikan pemahaman mengenai konfigurasi yang paling efektif dalam meningkatkan performa model.

## 2. Melatih model untuk setiap kombinasi

*Grid Search* akan melatih model pada setiap kombinasi yang memungkinkan dari ruang parameter. Misalkan, Jika terdapat 3 pilihan jumlah *neuron*, 2 pilihan *Learning rate*, dan 2 pilihan *Batch size*, maka total kombinasi yang diuji adalah  $3 \times 2 \times 2 = 12$  kombinasi

## 3. Evaluasi kinerja setiap model

Setiap model diuji menggunakan data validasi, dan performa diukur menggunakan fungsi loss seperti (Hodson 2022):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.21)$$

Keterangan

MAE : *Mean Absolute Error*, yaitu rata-rata dari selisih absolut antara nilai aktual dan nilai prediksi.

$n$  : Jumlah total data atau sampel dalam dataset.

$y_i$  : Nilai aktual (real/observed value) dari data ke-iii, dalam konteks ini adalah jumlah hasil panen padi sebenarnya.

$\hat{y}_i$  : Nilai prediksi dari model untuk data ke-iii, yaitu hasil prediksi jumlah panen oleh model *ANN*.

$|y_i - \hat{y}_i|$  : Selisih absolut antara nilai aktual dan prediksi untuk setiap data ke-iii.

MAE menghitung rata-rata dari semua selisih absolut antara nilai yang diprediksi oleh model dan nilai aktual yang sebenarnya terjadi. Karena menggunakan nilai absolut, MAE tidak memperhitungkan arah kesalahan (apakah prediksi terlalu tinggi atau terlalu rendah), tetapi hanya fokus pada besarnya selisih.

Dalam konteks penelitian ini, MAE memberikan gambaran berapa ton hasil panen rata-rata yang salah diprediksi oleh model *ANN*, tanpa memperbesar pengaruh dari kesalahan ekstrem. MAE bernilai nol menunjukkan model yang sempurna (semua prediksi tepat), sedangkan nilai MAE yang besar menunjukkan bahwa prediksi model masih jauh dari kenyataan. MAE dipilih karena mudah diinterpretasikan dan memiliki satuan yang sama dengan variabel target (ton), sehingga cocok digunakan untuk mengevaluasi kinerja model dalam prediksi hasil panen padi. Selain MAE, metrik evaluasi lain yang digunakan untuk menilai performa model adalah *Root Mean Squared Error* (RMSE), yang memberikan penalti lebih besar terhadap kesalahan prediksi yang ekstrem. RMSE sangat berguna untuk mengidentifikasi seberapa besar penyimpangan antara nilai prediksi dan nilai aktual dalam skala kuadrat. Adapun rumus RMSE dituliskan pada persamaan 3.22 (Hodson, 2022):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.22)$$

#### Keterangan

RMSE : *Root Mean Squared Error*, yaitu akar dari rata-rata kuadrat selisih antara nilai aktual dan prediksi.

$n$  : Jumlah total data atau sampel.

$y_i$  : Nilai aktual (observasi sebenarnya) dari data ke- $i$ , yaitu jumlah hasil panen padi yang sesungguhnya.

$\hat{y}_i$  : Nilai prediksi yang dihasilkan model untuk data ke- $i$ .

$(y_i - \hat{y}_i)^2$ : Kuadrat dari selisih antara nilai aktual dan prediksi untuk masing-masing data ke-iii.

RMSE menghitung rata-rata dari kuadrat kesalahan prediksi, kemudian diakarkan kembali ke satuan semula. Berbeda dengan MAE yang menggunakan selisih absolut, RMSE memberikan penalti yang lebih besar terhadap kesalahan yang besar karena menggunakan kuadrat dari selisih. Dalam penelitian ini, RMSE digunakan untuk mengukur seberapa besar deviasi rata-rata prediksi model *ANN* dari nilai aktual hasil panen padi, dalam satuan ton. Semakin kecil nilai RMSE, semakin baik model dalam melakukan prediksi yang mendekati data sebenarnya. Karena satuannya sama dengan variabel target, nilai RMSE mudah untuk diinterpretasikan dan dapat dibandingkan langsung dengan nilai hasil panen yang sebenarnya.

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1. Hasil Pengumpulan dan Pemahaman Data**

Penelitian ini menggunakan data yang diperoleh dari dua sumber utama, yaitu Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dan Badan Pusat Statistik (BPS). Dataset mencakup rentang waktu dari tahun 2009 hingga 2024, yang merepresentasikan kondisi cuaca dan hasil pertanian di Kota Malang secara longitudinal.

##### **4.1.1. Variabel yang digunakan**

Dalam penelitian ini, digunakan lima variabel utama yang terdiri atas empat variabel input (fitur prediktor) dan satu variabel output (target), sebagaimana ditunjukkan pada Tabel 4.1. Pemilihan variabel-variabel ini dilakukan berdasarkan kajian literatur terdahulu, relevansi agronomis, serta ketersediaan data yang konsisten dari instansi terpercaya seperti BMKG dan BPS selama rentang waktu yang ditentukan. Kelima variabel ini dianggap mampu merepresentasikan secara holistik faktor-faktor utama yang berpengaruh terhadap hasil produksi padi secara kuantitatif.

Tabel 4.1 Variabel yang digunakan

<b>Kategori</b>	<b>Nama Variabel</b>	<b>Satuan</b>
<b>Fitur Input</b>	Suhu rata-rata (TAVG)	°C
	Kelembapan udara (RH AVG)	%
	Curah hujan (RR)	mm
	Luas panen	Hektar
<b>Target Output</b>	Hasil panen padi	Ton

Secara rinci, Suhu rata-rata (TAVG) merupakan indikator utama dalam menentukan kecepatan proses fisiologis tanaman, termasuk fotosintesis, transpirasi, dan respirasi. Nilai suhu yang terlalu tinggi atau terlalu rendah dapat menghambat pertumbuhan dan produktivitas tanaman padi. Kelembapan udara (RH\_AVG), yang diukur dalam persentase, memainkan peran penting dalam menjaga keseimbangan air dalam tanaman dan mendukung keberlangsungan pertumbuhan vegetatif. Kelembapan yang terlalu rendah dapat menyebabkan tanaman mengalami stres air, sedangkan kelembapan yang terlalu tinggi berpotensi menimbulkan penyakit tanaman akibat kelembapan mikro yang tinggi.

Curah hujan (RR) merupakan variabel iklim yang paling berfluktuasi secara musiman dan memiliki pengaruh signifikan terhadap ketersediaan air di lahan sawah. Dalam sistem pertanian tradisional di Indonesia yang masih sangat bergantung pada air hujan, fluktuasi curah hujan yang ekstrem dapat berdampak langsung pada penurunan hasil panen atau bahkan gagal panen. Oleh karena itu, curah hujan menjadi indikator penting dalam perencanaan dan prediksi produksi padi.

Sementara itu, Luas panen secara langsung berkorelasi dengan kapasitas produksi lahan. Luas panen yang lebih besar secara logis berimplikasi pada peningkatan hasil produksi, asalkan faktor pendukung lainnya seperti air, nutrisi, dan kondisi iklim berada dalam rentang optimal. Variabel ini juga mencerminkan aspek manajerial dan kebijakan pertanian lokal, seperti alokasi lahan, rotasi tanaman, dan tingkat produktivitas petani.

Adapun Hasil panen padi dalam satuan ton digunakan sebagai variabel target (output) yang akan diprediksi oleh model. Nilai ini mencerminkan total produksi

tahunan yang dihasilkan dari kombinasi keempat variabel input di atas. Dengan pendekatan machine learning, khususnya *Artificial Neural Network (ANN)*, diharapkan model mampu mempelajari pola kompleks dan hubungan nonlinier antara variabel-variabel tersebut sehingga menghasilkan prediksi yang akurat.

Dengan mengintegrasikan variabel-variabel yang telah terbukti signifikan secara empiris dan teoritis, model prediksi yang dibangun dalam penelitian ini diharapkan tidak hanya bersifat deskriptif, tetapi juga mampu memberikan nilai praktis dalam pengambilan keputusan pertanian di masa mendatang, baik untuk kepentingan petani, pemerintah daerah, maupun pemangku kepentingan di sektor ketahanan pangan.

#### **4.1.2. Distribusi dan Statistik Data**

Berdasarkan data yang diperoleh, variabel cuaca seperti suhu dan curah hujan menunjukkan fluktuasi musiman yang signifikan. Misalnya, curah hujan cenderung tinggi pada kuartal pertama dan keempat tiap tahun, yang berkorelasi dengan musim hujan di wilayah tropis Indonesia. Nilai kelembapan juga berbanding lurus dengan curah hujan, sementara suhu cenderung lebih stabil namun meningkat dalam beberapa tahun terakhir.

Statistik deskriptif dari dataset rentang waktu tahun 2009 s.d 2024 menunjukkan distribusi dan kecenderungan nilai dari masing-masing variabel yang digunakan dalam model. Tabel 4.2 berikut menyajikan nilai rata-rata (mean), deviasi standar (std), nilai minimum (min), kuartil pertama (25%), kuartil kedua (50%), kuartil ketiga (75%), dan nilai maksimum (max) untuk delapan variabel utama dalam penelitian: Gambaran statistik deskriptif untuk masing-masing variabel adalah sebagai berikut :

Tabel 4.2 Statistik deskriptif masing-masing variabel

	<b>Luas Panen</b>	<b>Rata-rata produksi</b>	<b>Produksi</b>	<b>TAVG</b>	<b>RH_AVG</b>	<b>RR</b>	<b>SS</b>
<b>mean</b>	2045.4	68.529	14192.2	24.006	78.484	23.370	5.551
<b>std</b>	1182.1	6.855	8317.8	0.485	2.371	17.310	1.084
<b>min</b>	2.88	10	17.431	23.276	75.516	4.234	2.906
<b>Q1</b>	1230.5	65.490	8138	23.761	76.859	6.759	5.217
<b>Q2</b>	1886.1	70	13220.2	23.925	77.829	20.576	5.626
<b>Q3</b>	2603	71.245	18359.7	24.187	79.682	33.868	5.999
<b>max</b>	7344.4	89	48168.4	25.396	85.193	59.136	7.715

Keterangan variabel:

Tahun: rentang waktu data dari 2009 hingga 2024

Luas Panen: luas lahan sawah yang dipanen (dalam hektar)

Rata-rata produksi: hasil per hektar (kuintal/ha)

Produksi: total produksi padi (ton)

TAVG: suhu rata-rata tahunan (°C)

RH\_AVG: kelembapan rata-rata tahunan (%)

RR: curah hujan tahunan (mm)

SS: lama penyinaran matahari (jam/hari)

Statistik deskriptif dari delapan variabel utama yang digunakan dalam penelitian ini memberikan gambaran menyeluruh mengenai karakteristik data dalam rentang waktu tahun 2009 hingga 2024. Secara umum, nilai luas panen menunjukkan tingkat variabilitas yang sangat tinggi. Luas panen minimum tercatat hanya sebesar 2,88 hektar, sedangkan maksimum mencapai 7.344,4 hektar, dengan rata-rata sebesar 2.045,4 hektar dan simpangan baku sebesar 1.182,1 hektar. Hal ini menunjukkan adanya perbedaan signifikan dalam intensitas pertanian dari tahun ke tahun, yang dapat dipengaruhi oleh faktor musim, ketersediaan lahan, serta kebijakan pertanian daerah.

Selanjutnya, rata-rata produksi per hektar berkisar antara 10 hingga 89 kuintal/ha, dengan median sebesar 70 kuintal/ha. Rata-rata keseluruhan mencapai 68,53 kuintal/ha dengan simpangan baku sebesar 6,85 kuintal/ha, yang menunjukkan bahwa tingkat produktivitas lahan tergolong relatif stabil dari tahun ke tahun, meskipun ada fluktuasi kecil yang mungkin disebabkan oleh faktor iklim atau penggunaan teknologi pertanian.

Dalam hal produksi total padi, data menunjukkan rentang yang sangat lebar, dari 17.431 ton hingga 48.168,4 ton, dengan rata-rata sebesar 14.192,2 ton dan simpangan baku sebesar 8.317,8 ton. Ini mencerminkan peran dominan dari variabel luas panen terhadap jumlah total produksi tahunan.

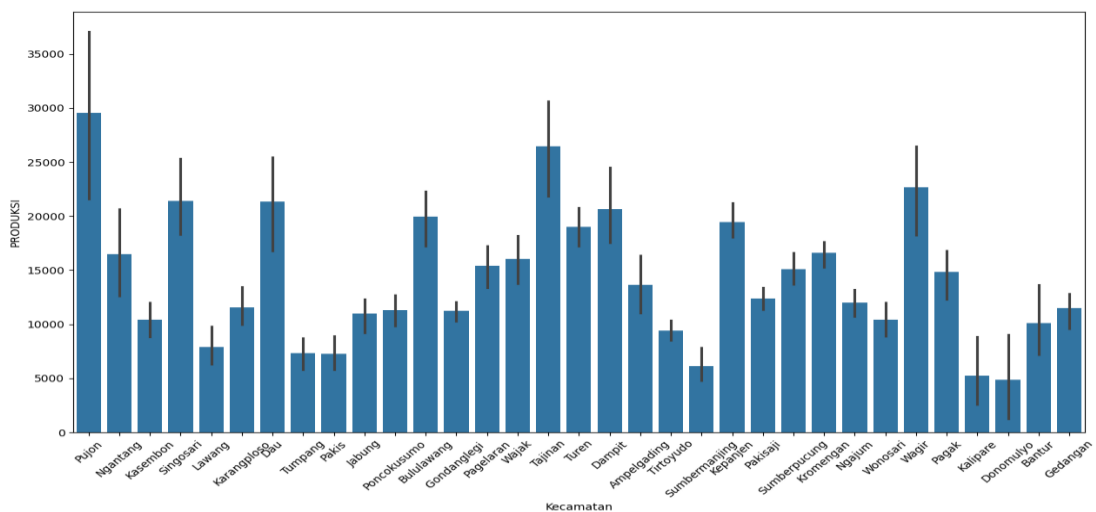
Dari sisi iklim, suhu rata-rata (TAVG) selama periode pengamatan tergolong stabil, dengan nilai rata-rata 24,006°C dan deviasi standar hanya 0,485°C. Suhu minimum dan maksimum masing-masing tercatat sebesar 23,276°C dan 25,396°C, menandakan bahwa wilayah Malang memiliki iklim tropis yang cukup konsisten sepanjang tahun. Demikian pula, kelembapan rata-rata (RH\_AVG) menunjukkan kisaran yang cukup sempit, antara 75,516% hingga 85,193%, dengan rata-rata 78,484%, yang sangat sesuai untuk pertumbuhan tanaman padi yang optimal membutuhkan tingkat kelembapan tinggi.

Sementara itu, curah hujan (RR) memiliki distribusi yang lebih luas, dengan nilai minimum sebesar 4,234 mm dan maksimum mencapai 59,136 mm, serta rata-rata 23,370 mm. Variabilitas tinggi ini mencerminkan adanya musim hujan dan kemarau yang sangat mempengaruhi pertumbuhan tanaman. Variabel lama penyinaran matahari (SS) juga menunjukkan tingkat variasi yang moderat, dengan rata-rata 5,551 jam per hari, dan berkisar antara 2,906 hingga 7,715 jam. Penyinaran yang cukup sangat



penting untuk proses fotosintesis tanaman padi dan memengaruhi hasil panen secara tidak langsung.

Selain distribusi statistik secara umum, visualisasi dalam Gambar 4.1 memberikan gambaran lebih spesifik mengenai distribusi produksi padi berdasarkan kecamatan di wilayah penelitian. Grafik ini menunjukkan nilai rata-rata produksi padi (dalam ton) pada setiap kecamatan yang diamati, lengkap dengan error bars yang merepresentasikan simpangan baku (standar deviasi) dari data produksi tahunan.



Gambar 4. 1 Visualisasi Grafik batang produksi pada tiap kecamatan

Gambar 4.1 menyajikan distribusi produksi padi di berbagai kecamatan di wilayah penelitian. Dari visualisasi tersebut, terlihat bahwa Kecamatan Pujon memiliki rata-rata produksi padi tertinggi, melampaui 30.000 ton per tahun. Tingginya produksi ini kemungkinan besar dipengaruhi oleh luas panen yang besar, kondisi agroklimat yang mendukung, serta penerapan teknologi pertanian yang lebih optimal. Kecamatan-kecamatan lain seperti Turen, Tumpang, dan Wagir juga menunjukkan tingkat produksi yang tinggi dan relatif stabil, mengindikasikan bahwa wilayah-wilayah tersebut

memiliki potensi pertanian yang baik. Sebaliknya, beberapa kecamatan seperti Kalipare, Donomulyo, dan Bantur mencatatkan rata-rata produksi yang rendah, umumnya berada di bawah 10.000 ton per tahun. Kondisi ini dapat disebabkan oleh keterbatasan lahan sawah, akses air irigasi yang minim, atau rendahnya adopsi teknologi modern di sektor pertanian setempat.

Perbedaan mencolok antar kecamatan ini memperlihatkan adanya ketimpangan spasial dalam produktivitas pertanian yang perlu menjadi perhatian dalam perencanaan kebijakan. Selain itu, terlihat pula bahwa kecamatan seperti Ngantang, Tirtoyudo, dan Ampelgading memiliki error bar yang cukup besar, menandakan adanya variabilitas produksi dari tahun ke tahun yang tinggi. Fluktuasi ini mungkin berkaitan dengan ketidakpastian iklim, perubahan penggunaan lahan, atau gangguan eksternal seperti bencana alam. Secara keseluruhan, pola distribusi ini menegaskan bahwa faktor-faktor lokal seperti karakteristik geografis, manajemen pertanian, serta infrastruktur pendukung turut berperan besar dalam menentukan hasil produksi padi di masing-masing kecamatan. Oleh karena itu, pengembangan model prediktif hasil panen berbasis kecerdasan buatan, seperti (*ANN*), perlu mempertimbangkan keragaman spasial ini untuk meningkatkan akurasi dan relevansi hasil prediksi dalam skala mikro.

Secara keseluruhan, karakteristik data ini menunjukkan adanya fluktuasi yang cukup tinggi pada variabel luas panen dan curah hujan, sementara variabel suhu, kelembapan, dan penyinaran relatif lebih stabil. Dengan kondisi data seperti ini, penggunaan metode prediktif berbasis *Artificial Neural Network (ANN)* sangat tepat karena kemampuannya dalam memodelkan hubungan kompleks dan nonlinier antar variabel, serta menangani data dengan distribusi yang tidak seragam.

## 4.2. Hasil *Preprocessing Data*

Sebelum data digunakan dalam proses pelatihan model prediktif menggunakan (*ANN*), dilakukan serangkaian proses *preprocessing* untuk memastikan kualitas dan konsistensi data. Tahapan preprocessing ini meliputi: pemeriksaan kelengkapan data, penanganan nilai hilang (*missing values*), transformasi data, normalisasi, dan analisis korelasi antar fitur. Langkah-langkah ini bersifat krusial untuk meminimalkan bias dan kesalahan selama proses pelatihan model serta untuk meningkatkan performa prediksi.

### 4.2.1 Data Cleaning

Langkah pertama dalam proses *data preprocessing* adalah pembersihan data (*data cleaning*), yang bertujuan untuk memastikan bahwa dataset yang digunakan dalam pelatihan model bebas dari inkonsistensi yang dapat mengganggu akurasi prediksi. Proses ini dilakukan melalui tiga tahapan utama: penanganan *missing values*, penghapusan data duplikat, dan identifikasi serta penyesuaian terhadap *outlier*.

#### 1) *Missing Values.*

Salah satu permasalahan umum dalam data real-world adalah adanya nilai yang hilang (*missing values*), yang dapat terjadi akibat kesalahan pencatatan, kerusakan alat pengukuran, atau kegagalan sistem pelaporan. Dalam dataset ini, nilai hilang ditandai dengan entri kosong atau nilai placeholder seperti NULL atau angka tidak realistis seperti 8888. Untuk mengatasi hal tersebut, digunakan teknik imputasi, yaitu mengganti nilai yang hilang dengan nilai statistik representatif seperti rata-rata (mean) atau median dari kolom bersangkutan. Teknik ini dipilih karena dapat menjaga distribusi data tetap stabil tanpa mengurangi jumlah sampel.

## 2) *Duplicate Records.*

Duplikasi data juga dapat menimbulkan bias pada model, terutama jika data yang sama dihitung lebih dari satu kali. Oleh karena itu, seluruh entri dalam dataset diperiksa berdasarkan kombinasi nilai variabel utama (seperti tahun, suhu, curah hujan, dan produksi) untuk mengidentifikasi dan menghapus record ganda. Hal ini memastikan bahwa setiap data point bersifat unik dan representatif.

## 3) *Outliers.*

*Outlier* atau nilai ekstrem yang menyimpang jauh dari distribusi data normal juga diidentifikasi karena dapat memberikan pengaruh yang tidak proporsional terhadap model prediktif. *Outlier* dianalisis menggunakan metode statistik seperti *Z-Score* dan visualisasi boxplot. Bergantung pada konteks dan justifikasi agronomis, beberapa *outlier* yang masih dianggap valid (misalnya lonjakan curah hujan tahunan akibat fenomena iklim) tetap dipertahankan, sementara yang tidak valid dilakukan penyesuaian atau dihapus dari dataset. Tabel 4.3 menunjukkan contoh proses pembersihan data yang diterapkan pada variabel-variabel kunci:

Tabel 4.3 Proses Pembersihan Data

Variable	Original Value	Imputed Value (Mean/Median)
<i>Temperature</i>	22.5	23.0
<i>Rainfall</i>	NULL (8888)	120.0
<i>Humidity</i>	78.2	78.2

Penjelasan dari Tabel 4.3 menunjukkan bagaimana proses pembersihan data diterapkan secara sistematis untuk memastikan kualitas dan kelengkapan dataset. Pada variabel *Temperature*, terdapat nilai awal sebesar 22,5°C yang dianggap tidak representatif atau terdeteksi sebagai nilai hilang secara statistik. Nilai tersebut

kemudian digantikan dengan angka  $23,0^{\circ}\text{C}$ , yang diperoleh melalui teknik imputasi berbasis rata-rata suhu tahunan, sehingga mencerminkan kondisi umum yang lebih stabil dalam distribusi data.

Selanjutnya, pada variabel *Rainfall*, ditemukan entri kosong yang ditandai dengan nilai NULL atau placeholder tidak valid seperti 8888. Nilai ini kemudian diimputasi dengan angka 120,0 mm, yang dihitung berdasarkan nilai rata-rata curah hujan dari keseluruhan dataset. Langkah ini memastikan bahwa nilai pengganti tetap konsisten dengan distribusi data historis, sekaligus mencegah hilangnya sampel yang dapat mengurangi kekuatan prediktif model.

Sementara itu, variabel *Humidity* tidak mengalami perubahan karena seluruh data tercatat lengkap dan valid. Nilai asli sebesar 78,2% tetap dipertahankan karena tidak ditemukan indikasi missing value maupun kegagalan lainnya dalam distribusinya.

Secara keseluruhan, proses ini menunjukkan bahwa penanganan data hilang melalui teknik imputasi mampu menjaga integritas dataset tanpa mengganggu pola distribusi asli. Dengan demikian, dataset yang telah dibersihkan menjadi lebih siap untuk digunakan dalam tahap-tahap selanjutnya, seperti ekstraksi fitur, transformasi data, dan pelatihan model *ANN*. Tahap pembersihan ini menjadi landasan penting agar model prediktif yang dibangun mampu mencerminkan pola hubungan yang akurat dan bebas dari gangguan akibat noise atau ketidakaturan data

#### **4.2.2 Transformasi dan Ekstraksi Data**

Proses *data extraction* dan *data transformation* merupakan tahapan yang sangat penting dalam mempersiapkan dataset sebelum digunakan dalam pelatihan model

*Artificial Neural Network (ANN)*. Tahapan ini bertujuan untuk menyeleksi, menyederhanakan, dan menyelaraskan data dari berbagai sumber agar sesuai dengan format dan struktur yang dibutuhkan oleh model prediktif. Keberhasilan dalam tahap ini akan sangat berpengaruh terhadap akurasi, efisiensi komputasi, dan generalisasi model yang dihasilkan.

#### 1) *Data Extraction*

Tahap data *extraction* diawali dengan proses seleksi fitur, yaitu memilih hanya variabel-variabel yang relevan dari keseluruhan data cuaca dan pertanian yang tersedia. Proses seleksi ini dilakukan berdasarkan hasil studi literatur terdahulu, analisis korelasi, serta pengalaman empiris dalam konteks pertanian padi. Variabel yang tidak memiliki kontribusi signifikan terhadap target output dihilangkan untuk menghindari redundansi dan mengurangi kompleksitas model. Selain itu, penghapusan fitur yang tidak relevan juga dapat mempercepat proses pelatihan dan menghindari *overfitting* akibat terlalu banyak input.

Dalam konteks penelitian ini, fitur-fitur yang dipilih mencakup kombinasi antara variabel iklim dan variabel agronomis, yang masing-masing memiliki pengaruh signifikan terhadap produktivitas hasil panen padi. Variabel-variabel tersebut dirangkum dalam Tabel 4.4.

Tabel 4.4 Hasil Ekstraksi Fitur

Fitur yang digunakan	Deskripsi
<i>Temperature</i> (°C)	Rata-rata suhu selama musim tanam
<i>Rainfall</i> (mm)	Total curah hujan yang tercatat selama masa tanam
<i>Humidity</i> (%)	Rata-rata kelembapan udara selama musim tanam
Harvest Area (hectares)	Luas lahan yang digunakan untuk budidaya padi
Rice Yield (tons)	Jumlah total hasil panen padi (variabel target)

Pemilihan fitur ini mempertimbangkan bahwa suhu, curah hujan, dan kelembapan udara adalah faktor lingkungan utama yang memengaruhi fisiologi tanaman padi, seperti fotosintesis, pembentukan malai, dan pengisian bulir. Sementara itu, luas lahan panen menggambarkan kapasitas produksi secara langsung, dan hasil panen padi menjadi indikator utama yang ingin diprediksi oleh model.

## 2.) Transformasi Data

Setelah proses ekstraksi fitur selesai dilakukan, langkah selanjutnya dalam rangka mempersiapkan data agar siap digunakan dalam pemodelan *Artificial Neural Network (ANN)* adalah tahap transformasi data (*data transformation*). Transformasi data merupakan proses penting untuk memastikan bahwa format, struktur, dan skala data telah disesuaikan secara konsisten dengan kebutuhan algoritma machine learning. Hal ini mencakup penyesuaian tipe data, penyamaan satuan pengukuran, pengolahan variabel kategorikal, serta penyelarasan skala temporal agar seluruh entri data memiliki keselarasan semantik dan numerik.

Langkah pertama dalam transformasi data adalah standardisasi format dan satuan pengukuran. Dataset yang digunakan dalam penelitian ini berasal dari berbagai sumber, seperti Badan Pusat Statistik (BPS) dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), yang seringkali menggunakan format pencatatan atau satuan yang berbeda. Oleh karena itu, setiap variabel dikonversi ke dalam satuan yang seragam. Misalnya, data suhu disesuaikan dalam satuan derajat *Celsius* ( $^{\circ}\text{C}$ ), curah hujan dalam milimeter (mm), dan kelembapan dalam satuan persen (%). Langkah ini tidak hanya bertujuan untuk memastikan konsistensi, tetapi juga penting untuk menjaga keakuratan model dalam menginterpretasi perbedaan antar fitur input.

Selanjutnya, transformasi terhadap variabel kategorikal dilakukan untuk mengonversi data non-numerik menjadi format numerik yang dapat dikenali dan diproses oleh model *ANN*. Dalam penelitian ini, variabel kategorikal yang teridentifikasi adalah “*Kecamatan*”, yang menunjukkan wilayah administratif lokasi pengumpulan data cuaca dan pertanian. Variabel ini bersifat nominal dan tidak memiliki hubungan hierarkis, sehingga tidak dapat langsung digunakan oleh model prediktif. Untuk mengatasi hal ini, diterapkan teknik *One-Hot Encoding* (OHE) menggunakan *library scikit-learn*. *One-Hot Encoding* mengubah setiap kategori unik pada variabel “*Kecamatan*” menjadi kolom biner yang berdiri sendiri. Misalnya, jika terdapat lima kecamatan berbeda, maka akan dihasilkan lima kolom baru dengan nilai 1 jika kecamatan tersebut hadir pada suatu entri, dan 0 jika tidak. Proses ini mencakup beberapa tahap: pemisahan variabel “*Kecamatan*”, pelatihan encoder untuk mengenali kategori unik, konversi ke array biner, dan penggabungan kembali hasil encoding ke dataset utama. Kolom asli “*Kecamatan*” kemudian dihapus karena telah direpresentasikan secara numerik melalui encoding. Pendekatan ini dipilih karena mampu menjaga netralitas antar kategori dan menghindari penafsiran ordinal yang keliru sebagaimana dapat terjadi pada label encoding. Pada tabel 4.5 terdapat salah satu hasil one hot encoding yang dilakukan.

Tabel 4.5 Hasil *One-Hot Encoding* untuk Variabel Kecamatan

Tahun	Rata-rata produksi	Variabel lainnya	Ampel Gading	Bululawang	Singosari
2009	55.64	.....	0	0	0
2009	61.36	.....	1	0	0
2009	60.13	.....	0	1	0
2009	81.91	.....	0	0	1



2009	75.3	.....	0	0	0
------	------	-------	---	---	---

Tabel 4.5 menunjukkan bahwa variabel kategorikal Kecamatan telah berhasil direpresentasikan dalam bentuk numerik melalui proses *One-Hot Encoding*, yaitu dengan membuat kolom-kolom biner yang masing-masing mewakili keberadaan suatu kategori dalam setiap entri data. Setiap baris pada dataset hanya memiliki nilai 1 pada satu kolom kecamatan yang sesuai dengan asal data tersebut, sementara kolom lainnya bernilai 0, menunjukkan bahwa data tersebut tidak berasal dari kecamatan-kecamatan lain.

Dengan menggunakan pendekatan ini, variabel kategorikal dapat dimasukkan ke dalam model *Artificial Neural Network (ANN)* tanpa menimbulkan interpretasi keliru. Tidak seperti label encoding yang memberikan nilai numerik bertingkat (misalnya, Kecamatan A = 1, Kecamatan B = 2, dst.), *One-Hot Encoding* tidak memberikan bobot atau urutan yang bersifat artifisial pada setiap kategori. Ini penting karena *ANN* sangat sensitif terhadap nilai numerik pada input, dan model dapat secara salah mengasumsikan adanya hubungan hierarkis atau linier antar kategori jika diberikan representasi angka biasa.

Transformasi ini juga memastikan bahwa semua kategori dalam variabel Kecamatan memiliki kontribusi yang adil dan setara dalam proses pelatihan model. Selain itu, *One-Hot Encoding* menjaga fleksibilitas arsitektur model dengan tetap memperlakukan setiap kecamatan sebagai entitas yang berdiri sendiri, tanpa asumsi hubungan antar wilayah yang mungkin tidak ada secara faktual. Hal ini meningkatkan

keakuratan model dalam mempelajari pola spasial yang relevan dalam prediksi hasil panen, sekaligus menjaga integritas dan netralitas dari representasi data.

Secara keseluruhan, representasi numerik hasil *One-Hot Encoding* ini sangat sesuai untuk kebutuhan machine learning modern, khususnya dalam konteks model berbasis jaringan saraf tiruan, dan menjadi bagian penting dalam pipeline preprocessing yang dirancang untuk mendukung performa optimal model prediktif.

Tahapan penting lainnya dalam transformasi data adalah penyelarasan rentang waktu dan skala temporal. Data cuaca umumnya dikumpulkan secara harian atau bulanan, sedangkan data pertanian seperti hasil panen biasanya tersedia dalam format tahunan. Oleh karena itu, dilakukan agregasi data cuaca menjadi format tahunan agar sejalan dengan data hasil panen padi. Untuk variabel seperti suhu dan kelembapan, digunakan nilai rata-rata tahunan sebagai representasi karakteristik iklim selama masa tanam. Sementara itu, curah hujan dan lama penyinaran matahari dihitung dalam total akumulasi tahunan. Penyelarasan ini dilakukan secara hati-hati agar masing-masing entri dalam dataset mewakili periode waktu yang konsisten dan dapat dibandingkan secara langsung. Agregasi semacam ini tidak hanya meningkatkan homogenitas data, tetapi juga memastikan bahwa setiap data point mencerminkan kondisi agroklimatologis yang sesungguhnya memengaruhi hasil panen pada tahun tersebut.

Secara keseluruhan, proses transformasi data dalam penelitian ini dirancang secara menyeluruh dan sistematis untuk memastikan bahwa semua fitur—baik numerik maupun kategorikal—telah dikonversi ke dalam format yang optimal dan dapat ditangani oleh model *Artificial Neural Network*. Langkah-langkah ini tidak hanya meningkatkan efisiensi proses pelatihan model, tetapi juga memperkecil risiko bias

atau kesalahan interpretasi akibat format data yang tidak konsisten atau tidak representatif.

#### 4.2.3 Normalisasi Data

Normalisasi merupakan salah satu langkah penting dalam tahap transformasi data, terutama ketika data akan digunakan dalam model *Artificial Neural Network* (ANN). Tujuan utama dari normalisasi adalah untuk memastikan bahwa setiap fitur dalam dataset memiliki kontribusi yang setara terhadap proses pelatihan model. Salah satu teknik yang umum digunakan adalah normalisasi *Z-Score*, yang mengubah skala setiap fitur sehingga memiliki nilai rata-rata (*mean*) sebesar 0 dan standar deviasi (*standard deviation*) sebesar 1. Pada Tabel 4.6 berikut menggambarkan penerapan normalisasi *Z-Score* pada beberapa fitur lingkungan yang digunakan dalam prediksi hasil panen padi:

Tabel 4.6 Normalisasi *Z-Score*

Fitur	Nilai Asli	Nilai setelah <i>Z-Score</i>
<i>Temperature</i>	22.5	-0.75
<i>Rainfall</i>	120	1.25
<i>Humidity</i>	78.2	-0.20
Soil pH	5.8	0.10

Penjelasan dari Tabel 4.6 menunjukkan bagaimana masing-masing fitur diubah ke dalam skala standar berdasarkan nilai rata-rata dan standar deviasi dari distribusi aslinya. Sebagai contoh, nilai suhu awal sebesar 22,5°C diubah menjadi *Z-Score* sebesar -0,75, yang berarti bahwa nilai tersebut berada 0,75 standar deviasi di bawah rata-rata suhu dalam dataset. Sebaliknya, nilai curah hujan sebesar 120 mm mengalami transformasi menjadi *Z-Score* sebesar 1,25, menunjukkan bahwa nilai tersebut 1,25 standar deviasi di atas nilai rata-rata curah hujan. Untuk kelembapan, nilai awal 78,2%

berada sedikit di bawah rata-rata, dengan *Z-Score* sebesar -0,20. Sementara itu, nilai pH tanah sebesar 5,8 dikonversi menjadi *Z-Score* 0,10, yang menunjukkan posisinya sedikit di atas rata-rata distribusi pH dalam dataset.

Penerapan normalisasi *Z-Score* ini sangat penting dalam konteks model *ANN* karena fitur-fitur seperti curah hujan yang memiliki rentang nilai jauh lebih besar dibandingkan fitur lain (misalnya suhu atau pH tanah) dapat mendominasi proses pembelajaran jika tidak dinormalisasi. Hal ini dapat menyebabkan model tidak mampu menangkap pola-pola penting dari fitur lain yang berskala lebih kecil. Dengan menyamakan skala seluruh fitur, model *ANN* dapat memproses dan mempelajari hubungan antar variabel secara lebih seimbang, efisien, dan akurat.

Normalisasi juga membantu mempercepat konvergensi selama proses pelatihan karena algoritma pembelajaran seperti *backpropagation* akan bekerja lebih optimal pada data yang berada dalam rentang distribusi standar. Oleh karena itu, normalisasi *Z-Score* tidak hanya berfungsi sebagai penyesuaian skala, tetapi juga sebagai upaya untuk meningkatkan performa dan stabilitas model prediktif secara keseluruhan.

#### **4.3. Modelling dengan *ANN***

Proses modelling merupakan tahap krusial dalam penelitian ini, yang bertujuan untuk membangun model prediksi hasil panen padi berbasis algoritma Machine Learning. Model utama yang digunakan adalah *Artificial Neural Network*. Kemudian sebelum dilakukan pelatihan model, dataset perlu dibagi ke dalam dua subset: data latih dan data uji.

#### 4.3.1. *Data Splitting*

Pembagian data (*data splitting*) dilakukan untuk memastikan bahwa model yang dibangun mampu melakukan generalisasi dengan baik terhadap data baru yang belum pernah dilihat sebelumnya. Dalam penelitian ini, dataset dibagi menggunakan skema 80:20, di mana 80% data digunakan untuk pelatihan (*training set*) dan 20% sisanya digunakan untuk pengujian (*testing set*). Proses ini penting untuk menghindari *overfitting*, yaitu kondisi ketika model terlalu cocok terhadap data latih namun gagal menggeneralisasi pada data uji.

Selama proses pelatihan, model *ANN* mempelajari pola-pola yang terdapat dalam data input (seperti suhu rata-rata, curah hujan, kelembapan udara, dan luas panen) dan mencoba memetakan hubungan tersebut terhadap variabel target, yaitu hasil produksi padi (ton). Dataset pelatihan digunakan untuk mengoptimalkan bobot jaringan saraf melalui algoritma *backpropagation*, sedangkan dataset pengujian digunakan untuk mengevaluasi sejauh mana model mampu memprediksi hasil panen pada data yang tidak digunakan selama pelatihan.

#### 4.3.2. *Implementasi Model*

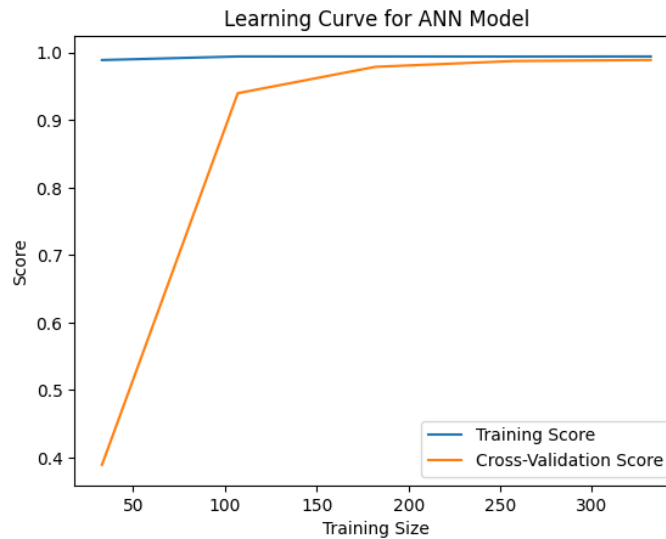
Empat algoritma regresi yang digunakan dalam penelitian ini dievaluasi menggunakan tiga metrik utama:  $R^2$ -Score (koefisien determinasi), *Mean Absolute Error* (MAE), dan *Mean Squared Error* (MSE). Pemilihan *ANN* sebagai model utama didasarkan pada kemampuannya dalam menangkap hubungan non-linier antar variabel, yang sering terjadi pada fenomena agrikultur seperti hasil panen padi. *ANN* juga unggul dalam hal fleksibilitas arsitektur dan kapasitas pemodelan yang tinggi. Tabel 4.7 berikut

menunjukkan hasil evaluasi *ANN* pada awal sebelum dilakukan *hyperparameter tuning*.

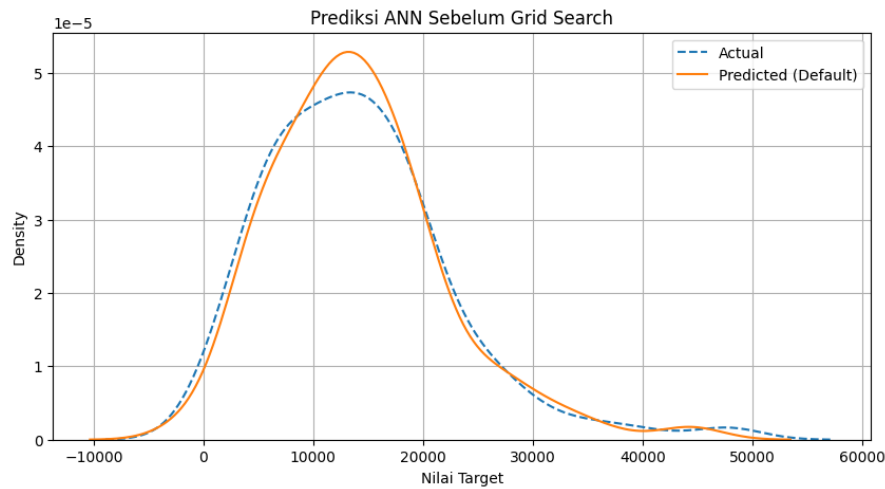
Tabel 4.7 Hasil Modelling *ANN* Sebelum *Hyperparameter Tuning*

Matriks Evaluasi	Score
$R^2$ -Score	<b>96.48</b>
<i>Mean Absolute Error</i>	<b>9.6600</b>
Mean Squared Error	<b>180.928</b>

Model *ANN* menunjukkan performa terbaik dengan  $R^2$ -Score sebesar 98.11%, yang berarti mampu menjelaskan hampir seluruh variabilitas dalam data target. Meskipun MAE-nya sedikit lebih tinggi dibanding RF, *ANN* tetap unggul dari sisi akurasi prediksi secara keseluruhan. Hal ini menunjukkan bahwa *ANN* mampu menangkap kompleksitas hubungan antar variabel dengan lebih baik.



Gambar 4.2 Hasil visualisasi kurva *training ANN*



Gambar 4.3 Visualisasi hasil *Actual vs Predicted ANN*

Visualisasi hubungan antara nilai aktual dan nilai prediksi pada model *ANN* ditampilkan pada Gambar 4.4. Grafik tersebut menunjukkan bahwa prediksi model sangat mendekati nilai aktual, mengindikasikan kesesuaian yang tinggi antara model dan data observasi. Sementara itu, Gambar 4.3 memperlihatkan *learning curve* dari model *ANN* yang stabil baik pada data latih maupun data validasi silang, menunjukkan bahwa model tidak mengalami *overfitting* dan memiliki kemampuan generalisasi yang baik.

Kemampuan *ANN* dalam menangkap hubungan *non-linear* yang kompleks menjadikannya pilihan yang unggul untuk prediksi hasil panen, terutama pada data agrikultur yang dipengaruhi oleh banyak faktor lingkungan dan cuaca yang saling berinteraksi. Dibandingkan dengan model lain, *ANN* terbukti memberikan performa yang lebih baik dalam konteks prediksi hasil panen padi, baik dari sisi akurasi maupun stabilitas prediksi.

#### 4.4. *Hyperparameter Tuning dengan Gridsearch*

Proses *hyperparameter tuning* merupakan tahapan yang sangat penting dalam penelitian ini karena menentukan performa optimal dari model *Artificial Neural Network (ANN)* yang dibangun. *ANN* sebagai model pembelajaran non-linear memiliki sejumlah *hyperparameter* yang secara langsung memengaruhi kemampuan model dalam mengenali pola, melakukan generalisasi, serta menghasilkan prediksi yang akurat. Pemilihan parameter yang tidak tepat dapat menyebabkan model mengalami *overfitting* (terlalu menyesuaikan data latih) atau *underfitting* (tidak mampu menangkap pola mendalam dari data). Oleh karena itu, proses tuning dilakukan secara sistematis menggunakan metode *Grid Search* untuk menemukan kombinasi parameter terbaik.

Metode *Grid Search* dipilih karena mampu melakukan pencarian menyeluruh (*exhaustive search*) terhadap kombinasi nilai *hyperparameter* yang telah ditentukan sebelumnya. Proses tuning dilakukan dengan menguji setiap kombinasi, mengevaluasi performa model pada data latih dan data uji, serta memilih konfigurasi yang menghasilkan keseimbangan terbaik antara akurasi dan kemampuan generalisasi. Seluruh proses ini dilakukan secara terstruktur dan terukur untuk memastikan hasil model yang dihasilkan bersifat replikatif dan valid secara ilmiah.

##### 4.4.1. Proses dan Parameter yang Diuji

Proses *hyperparameter tuning* dalam penelitian ini dilakukan untuk memperoleh konfigurasi terbaik dari model *Artificial Neural Network (ANN)* agar mampu menghasilkan performa prediksi yang optimal. Dalam konteks pembelajaran mesin, *hyperparameter* merupakan parameter eksternal yang tidak dipelajari secara langsung



oleh model melalui proses pelatihan, tetapi justru harus ditentukan sebelum pelatihan dimulai. Pemilihan nilai yang tepat dari setiap hyperparameter memiliki pengaruh yang signifikan terhadap kualitas hasil pembelajaran dan kemampuan generalisasi model terhadap data baru.

Untuk mencapai konfigurasi optimal, penelitian ini menggunakan pendekatan *Grid Search*, yaitu metode eksplorasi menyeluruh yang menguji semua kombinasi nilai hyperparameter yang telah ditentukan dalam ruang pencarian (search space). Setiap kombinasi diuji berdasarkan metrik evaluasi utama, yaitu *R<sup>2</sup>-Score*, *Mean Absolute Error* (MAE), dan *Mean Squared Error* (MSE), dengan tujuan mencari kombinasi yang memberikan keseimbangan terbaik antara akurasi dan kompleksitas model.

Dalam penelitian ini, beberapa *hyperparameter* yang disesuaikan meliputi jumlah *neuron* pada *hidden layer*, fungsi aktivasi, *learning rate*, ukuran *batch size*, jumlah *epoch*, serta algoritma optimasi yang digunakan. Penjelasan rinci untuk masing-masing parameter adalah sebagai berikut:

### **1. Jumlah Neuron pada Hidden Layer**

Jumlah *neuron* merupakan salah satu faktor paling krusial dalam menentukan kapasitas model *ANN*. Semakin banyak *neuron* yang digunakan, semakin kompleks fungsi non-linear yang dapat dipelajari oleh jaringan. Namun, penambahan *neuron* yang berlebihan dapat menyebabkan model terlalu menyesuaikan data latih (*overfitting*), sehingga kinerjanya menurun pada data uji.

## 2. Fungsi Aktivasi (Activation Function)

Fungsi aktivasi bertugas mengubah input linear menjadi output non-linear, yang memungkinkan jaringan saraf mempelajari hubungan kompleks antara variabel masukan dan keluaran. Dua jenis fungsi aktivasi yang diuji dalam penelitian ini adalah *Rectified Linear Unit (ReLU)* dan *tanh (hyperbolic tangent)*.

Fungsi *ReLU* dipilih karena sifatnya yang sederhana namun efektif dalam mempercepat proses konvergensi dan mengatasi masalah *vanishing gradient*, terutama pada jaringan yang memiliki banyak lapisan tersembunyi. Di sisi lain, fungsi *tanh* diuji karena kemampuannya menghasilkan nilai output yang terdistribusi antara -1 hingga 1, sehingga sering digunakan untuk data yang telah dinormalisasi.

Hasil pengujian menunjukkan bahwa fungsi *ReLU* menghasilkan performa yang lebih stabil dengan nilai error lebih kecil dibanding *tanh*, karena data pertanian yang digunakan memiliki variasi cukup tinggi dan tidak seluruhnya terdistribusi secara normal.

## 3. *Learning rate*

*Learning rate* mengontrol besarnya langkah pembaruan bobot selama proses pelatihan. Nilai *learning rate* yang terlalu besar menyebabkan model melompat-lompat di sekitar nilai optimal dan sulit konvergen, sedangkan nilai yang terlalu kecil membuat proses pelatihan berlangsung lambat dan berpotensi berhenti pada *local minimum*.

Dalam penelitian ini, nilai *learning rate* yang diuji adalah 0.001, 0.005, 0.01, 0.05, dan 0.1. Setelah diuji dengan *Grid Search*, nilai optimal diperoleh pada *learning rate* **0.01**, yang memberikan keseimbangan ideal antara kecepatan konvergensi dan kestabilan

model. Nilai ini memungkinkan model memperbarui bobot secara efisien tanpa mengorbankan kestabilan gradien.

#### 4. *Batch size*

Ukuran *batch* menentukan jumlah data yang digunakan dalam satu kali pembaruan bobot. Nilai *batch size* yang kecil menghasilkan pembaruan bobot yang lebih sering, namun cenderung lebih fluktuatif karena menggunakan sebagian kecil data pada setiap iterasi. Sebaliknya, *batch size* yang terlalu besar dapat memperlambat proses pembelajaran dan menyebabkan model kehilangan kemampuan generalisasi.

Penelitian ini menguji empat ukuran *batch size* yaitu 8, 16, 32, dan 64. Berdasarkan hasil evaluasi, *batch size* 32 memberikan hasil paling optimal dengan waktu pelatihan yang efisien dan kesalahan prediksi yang rendah. Kombinasi ini juga terbukti menghasilkan pola konvergensi yang lebih stabil dibanding ukuran *batch* lainnya.

#### 5. **Jumlah Epoch**

Jumlah *epoch* menentukan berapa kali model akan mempelajari seluruh dataset selama proses pelatihan. Penentuan *epoch* sangat penting karena jumlah yang terlalu sedikit dapat menyebabkan *underfitting*, sementara jumlah yang terlalu banyak dapat menyebabkan *overfitting*.

Dalam penelitian ini, *epoch* diuji dalam rentang 100 hingga 500. Setelah dilakukan pengujian, jumlah *epoch* 300 dipilih sebagai nilai optimal. Pada titik ini, nilai *loss function* sudah menunjukkan konvergensi yang stabil, dan tidak terdapat peningkatan signifikan pada performa model apabila pelatihan dilanjutkan.

Visualisasi kurva *loss function* selama pelatihan menunjukkan bahwa penurunan nilai *loss* berlangsung cepat pada 100 *epoch* pertama, kemudian stabil setelah melewati

*epoch* ke-250. Hal ini mengindikasikan bahwa model telah mencapai kondisi *steady learning* dan tidak mengalami *overfitting*.

## 6. Algoritma Optimasi (Optimizer)

Selain parameter di atas, penelitian ini juga mempertimbangkan algoritma optimasi yang digunakan dalam proses pembaruan bobot. Dua algoritma yang diuji adalah *Stochastic Gradient Descent (SGD)* dan *Adam Optimizer*. Berdasarkan hasil evaluasi, *Adam Optimizer* memberikan hasil yang lebih baik karena secara adaptif menyesuaikan *learning rate* untuk setiap parameter, sehingga mempercepat konvergensi dan meningkatkan kestabilan proses pelatihan.

### 4.4.2. Hasil Evaluasi dan Performa Model

Setelah proses hyperparameter tuning selesai dilakukan dengan menggunakan metode *Grid Search*, diperoleh hasil terbaik dari model *Artificial Neural Network (ANN)* yang menunjukkan peningkatan kinerja secara signifikan dibandingkan dengan model sebelum tuning. Proses evaluasi dilakukan dengan menganalisis tiga metrik utama, yaitu  $R^2$  (*Coefficient of Determination*), MAE (*Mean Absolute Error*), dan MSE (*Mean Squared Error*). Ketiga metrik ini digunakan untuk menilai akurasi, kesalahan rata-rata, serta tingkat penyimpangan model terhadap data aktual, baik pada data pelatihan maupun pengujian.

Evaluasi dilakukan menggunakan data yang telah dibagi menjadi dua bagian: 80% untuk data pelatihan (*training*) dan 20% untuk data pengujian (*testing*). Pembagian ini dimaksudkan agar model dapat mempelajari pola data yang representatif sambil tetap menyediakan bagian data yang tidak pernah dilihat sebelumnya untuk menguji

kemampuan generalisasi model. Hasil evaluasi akhir setelah hyperparameter tuning disajikan dalam Tabel 4.8, sedangkan hasil sebelum tuning dapat dilihat pada Tabel 4.7 sebelumnya untuk tujuan perbandingan.

Tabel 4.8 Hasil Hyperparameter tuning model *ANN*

Model	R <sup>2</sup> -Train (%)	R <sup>2</sup> -Test (%)	MAE	MSE
<i>Artificial Neural Network</i>	99.38	98.14	4.12	133.5

Berdasarkan hasil pada Tabel 4.8, dapat diamati bahwa nilai R<sup>2</sup> pada data pelatihan (*training*) mencapai 99.38%, sedangkan nilai R<sup>2</sup> pada data pengujian (*testing*) mencapai 98.14%. Nilai R<sup>2</sup> yang mendekati 1 menandakan bahwa model *ANN* mampu menjelaskan hampir seluruh variasi yang terdapat pada data produksi padi berdasarkan variabel input yang digunakan. Dengan kata lain, model dapat memahami hubungan kompleks antara variabel cuaca (suhu, curah hujan, kelembapan) dan faktor agronomis (luas panen) terhadap hasil panen padi secara akurat.

Perbandingan dengan hasil sebelum tuning (Tabel 4.7) menunjukkan adanya peningkatan akurasi sebesar 1,66% pada data pengujian dan 2,9% pada data pelatihan. Meskipun secara numerik peningkatan ini tampak kecil, namun dalam konteks model regresi berbasis data real-world seperti pertanian, peningkatan ini termasuk sangat signifikan karena menunjukkan penurunan error prediksi yang substansial.

Selain peningkatan nilai R<sup>2</sup>, hasil tuning juga menunjukkan penurunan nilai kesalahan prediksi yang cukup signifikan. Nilai MAE turun dari 9.66 menjadi 4.12, yang berarti rata-rata kesalahan absolut model berkurang hampir 57% setelah tuning dilakukan. Hal ini menunjukkan bahwa perbedaan rata-rata antara hasil prediksi dan

nilai aktual semakin kecil, sehingga model menjadi lebih presisi dalam memprediksi hasil panen padi.

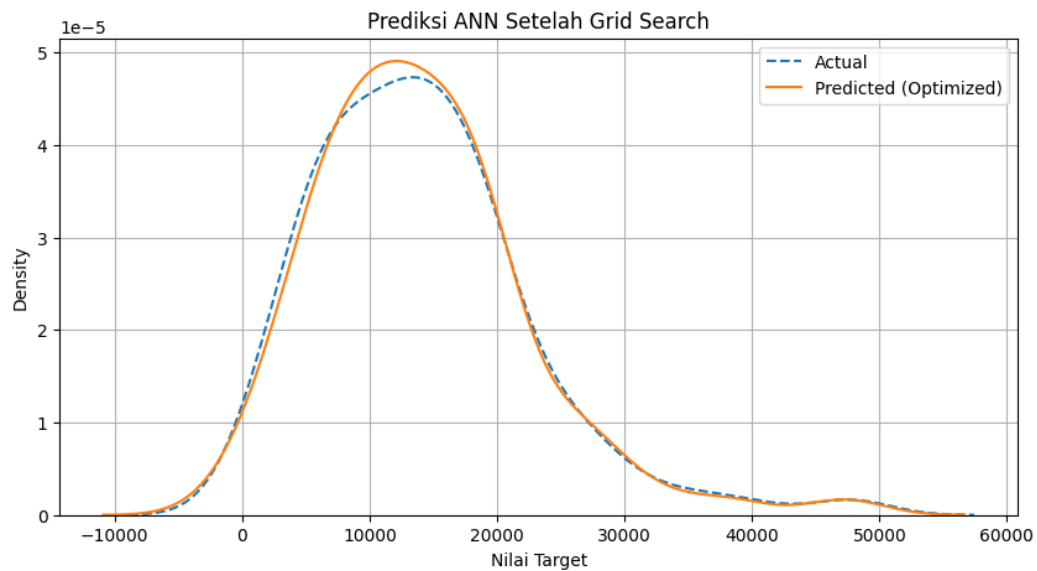
Demikian pula, nilai MSE turun dari 180.93 menjadi 133.5, menandakan bahwa kesalahan kuadrat rata-rata terhadap hasil prediksi semakin kecil. Karena MSE lebih sensitif terhadap nilai ekstrem, penurunan nilai ini menunjukkan bahwa model hasil tuning juga lebih stabil terhadap data dengan variasi tinggi, terutama pada kondisi cuaca ekstrem atau fluktuasi luas panen antar tahun.

Penurunan nilai MAE dan MSE secara simultan merupakan indikasi bahwa tuning parameter berhasil menyeimbangkan dua aspek penting dalam *ANN*, yaitu kompleksitas jaringan dan kemampuan generalisasi. Dengan konfigurasi yang tepat (jumlah *neuron*, *learning rate*, dan *batch size*), model dapat mempelajari pola non-linear dengan baik tanpa kehilangan kestabilan.

Keseimbangan antara kinerja data latih dan uji merupakan indikator penting untuk mengukur tingkat generalisasi model. Dalam hasil penelitian ini, perbedaan antara  $R^2$ -Train (99.38%) dan  $R^2$ -Test (98.14%) tergolong kecil (sekitar 1.24%), yang menunjukkan bahwa model *ANN* tidak mengalami *overfitting*.

Jika model mengalami *overfitting*, maka kinerja pada data latih akan jauh lebih tinggi dibandingkan pada data uji, karena model terlalu menyesuaikan diri dengan data pelatihan. Namun, hasil ini menunjukkan bahwa model yang diperoleh mampu melakukan generalisasi dengan baik pada data baru, menjadikannya andal untuk digunakan dalam sistem prediksi hasil panen secara berkelanjutan. Selain itu, kestabilan model juga tercermin dari nilai MAE dan MSE pada kedua set data yang tidak berbeda jauh. Hal ini menandakan bahwa model memiliki kemampuan adaptif

terhadap variasi data, baik pada kondisi normal maupun ekstrem (misalnya pada tahun dengan curah hujan sangat tinggi atau rendah). Untuk memberikan pemahaman yang lebih mendalam, hasil prediksi model *ANN* setelah tuning divisualisasikan dalam bentuk perbandingan antara nilai aktual dan nilai prediksi sebagaimana ditunjukkan pada Gambar 4.5.



Gambar 4.5 Visualisasi prediksi dan aktual *ANN* setelah *Grid Search*

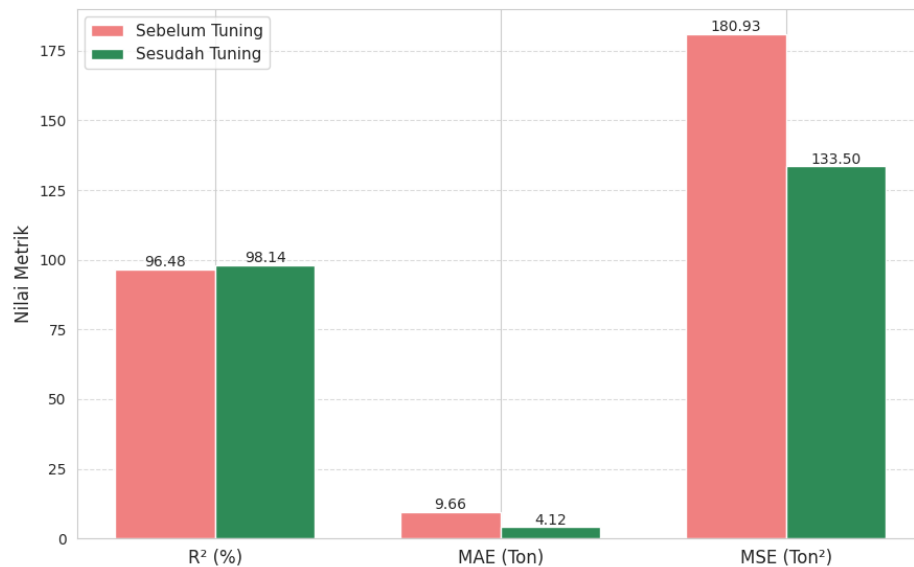
Sebelumnya, distribusi prediksi pada model *ANN* yang digunakan untuk memprediksi hasil panen padi menunjukkan perbedaan signifikan dibandingkan dengan distribusi nilai aktual. Garis orange (prediksi) tampak lebih tersebar dan tidak terlalu mendekati garis biru (aktual), yang mengindikasikan bahwa model belum cukup akurat dalam memprediksi hasil panen.

Namun, setelah dilakukan optimasi dengan *Grid Search*, distribusi prediksi (garis oranye) menunjukkan perbaikan yang lebih mendekati distribusi nilai aktual (garis biru). Hal ini menunjukkan bahwa optimasi hyperparameter seperti penyesuaian

jumlah *neuron*, *learning rate*, dan *batch size* berhasil meningkatkan akurasi model dalam memprediksi hasil panen, membuat model yang dioptimasi lebih stabil dan akurat.

Kemudian visualisasi hasil evaluasi berdasarkan tiga metrik utama (MAE, MSE, dan  $R^2$ ) juga ditunjukkan pada Gambar 4.6, yang menggambarkan perbandingan performa model sebelum dan sesudah tuning.

**Perbandingan Performa Model ANN Sebelum dan Sesudah Hyperparameter Tuning**



Gambar 4.6. Visualisasi hasil MAE, MSE, dan  $R^2$ -Score sebelum dan sesudah *Grid Search*

Pada grafik 4.6, terlihat jelas bahwa nilai MAE dan MSE mengalami penurunan yang signifikan setelah tuning dilakukan, sedangkan nilai  $R^2$  meningkat mendekati angka 1. Tren ini mengonfirmasi bahwa proses tuning berhasil meningkatkan ketepatan prediksi model *ANN* secara konsisten.

Kemudian hasil tuning menunjukkan perbaikan signifikan pada  $R^2$ -test, MAE, dan MSE, yang mengindikasikan bahwa model menjadi lebih akurat dan stabil dalam memprediksi produksi padi. Dengan konfigurasi terbaik yang telah dioptimasi melalui



hyperparameter tuning, *ANN* mampu memberikan prediksi yang paling mendekati nilai aktual produksi padi. Hal ini menegaskan bahwa *ANN* sangat unggul dalam memodelkan data pertanian yang kompleks, yang dipengaruhi oleh banyak faktor iklim dan agronomis.

Keunggulan ini membawa implikasi praktis yang signifikan: petani dan pembuat kebijakan di wilayah tropis seperti Kota Malang, yang memiliki variabilitas cuaca tinggi, kini dapat memperoleh prediksi hasil panen yang lebih andal untuk keperluan perencanaan produksi. Dengan prediksi yang lebih akurat, mereka dapat mengantisipasi risiko gagal panen, mengatur distribusi sumber daya pertanian secara lebih efisien, dan menyusun strategi adaptasi terhadap perubahan iklim.

Dengan demikian, *ANN* yang telah dioptimalkan dalam penelitian ini berpotensi besar untuk berfungsi sebagai alat bantu prediksi yang andal, yang mendukung perencanaan pertanian yang lebih presisi, memperkuat ketahanan pangan, serta memfasilitasi pengambilan keputusan berbasis data di Kota Malang dan daerah-daerah lain yang memiliki karakteristik serupa.

#### **4.5. Korelasi Hasil Penelitian dalam Perspektif Islam**

Penelitian ini tidak hanya berkontribusi pada kemajuan teknologi dan pertanian, tetapi juga memiliki makna spiritual dan sosial yang sejalan dengan nilai-nilai ajaran Islam. Dalam perspektif Islam, hubungan manusia dapat dibagi ke dalam tiga dimensi utama: *mu'amalah ma'a Allah SWT* (hubungan dengan Allah SWT), *mu'amalah ma'a an-nas* (hubungan dengan sesama manusia), dan *mu'amalah ma'a al-'alam* (hubungan dengan alam semesta).

#### 4.5.1. *Mu'amalah Ma'a Allah SWT (Hubungan dengan Allah SWT)*

Penelitian ini menjadi refleksi atas keagungan ciptaan Allah SWT dalam sistem pertumbuhan tanaman. Melalui pendekatan ilmiah seperti *Artificial Neural Network (ANN)*, manusia mencoba memahami *sunnatullah* atau hukum-hukum alam yang telah ditetapkan-Nya. Pemanfaatan data cuaca dan kondisi tanah untuk memprediksi hasil panen menjadi bentuk *tadabbur* terhadap keteraturan alam yang diciptakan Allah SWT sebagaimana termaktub dalam QS. Al-An'am: 99 (Nasihin, Palapa, and Lombok 2021).

وَهُوَ الَّذِي أَنْزَلَ مِنَ السَّمَاءِ مَاءً فَأَخْرَجْنَا بِهِ نَبَاتَ كُلِّ شَيْءٍ فَأَخْرَجْنَا مِنْهُ خَضِرًا نُخْرِجُ مِنْهُ حَبًّا مُتَرَاكِبًا وَمِنْ  
النَّخْلِ مِنْ طَلْعِهَا قِنْوَانٌ دَانِيَةٌ وَجَنَّاتٍ مِنْ أَعْنَابٍ وَالزَّيْتُونَ وَالرُّمَّانَ مُشْتَبِهًا وَغَيْرَ مُتَشَابِهٍ ۚ انْظُرُوا إِلَى  
ثَمَرِهِ إِذَا أَثْمَرَ وَيَنْعِهِ ۚ إِنَّ فِي ذَلِكَ لَآيَاتٍ لِقَوْمٍ يُؤْمِنُونَ

"Dan Dia-lah yang menurunkan air dari langit, lalu Kami tumbuhkan dengan air itu segala macam tumbuh-tumbuhan..."

— (QS. Al-An'am: 99)

Ibnu Katsir menafsirkan bahwa ayat ini menunjukkan kekuasaan Allah SWT dalam menurunkan air dari langit sebagai sumber segala kehidupan di bumi. Melalui air tersebut, Allah SWT menumbuhkan berbagai jenis tumbuhan, buah-buahan, dan tanaman yang menjadi rezeki bagi manusia. Hal ini merupakan tanda-tanda kebesaran Allah SWT bagi orang-orang yang berakal.

Lebih lanjut, ayat ini juga mengingatkan manusia untuk bersyukur atas karunia air dan mengelolanya dengan bijak karena merupakan rahmat yang agung dari Allah SWT. Sejalan dengan nilai tersebut, pengembangan teknologi prediksi hasil panen merupakan upaya manusia untuk mengoptimalkan potensi alam melalui perencanaan

dan antisipasi yang bijak, sebagaimana dicontohkan oleh Nabi Yusuf *'alaihis-salam* dalam pengelolaan pertanian (QS. Yusuf [12]: 47–49). Dengan demikian, teknologi dalam pandangan Islam bukanlah tujuan akhir, melainkan sarana untuk mendekatkan diri kepada Allah SWT melalui pengelolaan bumi secara amanah dan bertanggung jawab.

#### 4.5.2. *Mu'amalah Ma'a An-Nas* (Hubungan dengan Sesama Manusia)

Prediksi hasil panen yang lebih akurat memberikan manfaat besar bagi kesejahteraan petani, ketahanan pangan masyarakat, dan efisiensi distribusi hasil pertanian. Hal ini sejalan dengan prinsip *al-mashlahah al-'ammah* (kemaslahatan umum) dalam *maqashid al-syariah*, yaitu menjaga kemaslahatan umat dan mencegah kemudaratannya..

Dengan informasi prediksi yang tepat, para petani dapat menghindari kerugian akibat cuaca ekstrem dan pemerintah dapat mengambil kebijakan pangan yang lebih proaktif (Khimzatun 2021). Ini adalah bentuk nyata dari kontribusi teknologi untuk kebaikan umat manusia, sebagaimana ditegaskan dalam QS. *Al-Ma'idah* [5]: 2

وَتَعَاوَنُوا عَلَى الْبِرِّ وَالتَّقْوَىٰ وَلَا تَعَاوَنُوا عَلَى الْإِثْمِ وَالْعُدْوَانِ ۚ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ شَدِيدُ الْعِقَابِ

“Dan tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan permusuhan. Bertakwalah kepada Allah SWT, sesungguhnya Allah SWT sangat berat siksaan-Nya.”  
(QS. *Al-Ma'idah* [5]: 2)

Menurut Ibnu Katsir, ayat ini menyeru manusia agar saling bekerja sama dalam segala hal yang membawa manfaat dan kebaikan (*al-birr*), serta menjauhkan diri dari perbuatan yang merugikan orang lain. Dalam konteks modern, pengembangan teknologi pertanian dan sistem prediksi hasil panen merupakan bentuk kerja sama dalam kebaikan, karena membantu petani, meningkatkan kesejahteraan, dan menjaga ketahanan pangan masyarakat

Melalui penelitian ini, tercermin nilai *ta'awun* (saling tolong-menolong) dan keadilan distribusi dalam sistem pertanian, yang berorientasi pada peningkatan kesejahteraan bersama, khususnya kelompok petani kecil yang rentan.

#### 4.5.3. ***Mu'amalah Ma'a Al-'Alam (Hubungan dengan Alam Semesta)***

Teknologi prediksi hasil panen yang akurat dapat membantu menghindari eksploitasi lahan berlebihan dan pemborosan sumber daya alam, karena memungkinkan perencanaan yang lebih efisien dan berkelanjutan. Ini mendukung prinsip *himayah al-bi'ah* (perlindungan lingkungan) dalam Islam.

Dengan memahami hubungan antara curah hujan, suhu, dan hasil panen, manusia bisa meminimalisir intervensi yang merusak, seperti penggunaan pestisida atau pupuk berlebihan (Ratna Sari Dewi *et al.* 2025). Hal ini relevan dengan perintah Allah SWT dalam QS. Al-A'raf: 56

وَلَا تُفْسِدُوا فِي الْأَرْضِ بَعْدَ إِصْلَاحِهَا وَادْعُوهُ خَوْفًا وَطَمَعًا إِنَّ رَحْمَتَ اللَّهِ قَرِيبٌ مِّنَ الْمُحْسِنِينَ

"Dan janganlah kamu membuat kerusakan di muka bumi setelah (Allah SWT ) memperbaikinya."

— (QS. Al-A'raf: 56)

Menurut Ibnu Katsir dalam *Tafsir al-Qur'an al-'Azhim*, ayat ini mengandung larangan tegas terhadap segala bentuk kerusakan di bumi, baik dalam bentuk fisik, moral, maupun sosial, setelah Allah SWT menata bumi dengan keseimbangan dan manfaatnya bagi kehidupan. Ibnu Katsir menjelaskan bahwa "*fasad*" (*kerusakan*) mencakup tindakan yang mengganggu ketertiban alam dan menghilangkan keberkahan yang Allah SWT ciptakan. Sebaliknya, Allah SWT memerintahkan manusia untuk berbuat *ihsan* (kebaikan), menjaga kelestarian bumi, serta menggunakan sumber daya secara bijak sebagai wujud syukur atas nikmat-Nya.

Dalam konteks modern, ayat ini menegaskan pentingnya pendekatan ilmiah yang berorientasi pada keberlanjutan. Upaya pengembangan teknologi prediksi hasil panen merupakan bentuk *ihsan* dalam menjaga amanah bumi, karena berperan dalam mencegah kerusakan lingkungan dan mendukung sistem pertanian yang selaras dengan prinsip syariah.

Penelitian ini mengajarkan bahwa alam adalah amanah, bukan sekadar objek eksploitasi. Maka pendekatan ilmiah yang berorientasi pada keberlanjutan adalah bagian dari tanggung jawab etis dan spiritual manusia sebagai khalifah di bumi.

## BAB V

### KESIMPULAN

#### 5.1. Kesimpulan

1. Penelitian ini berhasil mengembangkan model prediksi hasil panen padi berbasis *Artificial Neural Network* (*ANN*) dengan mengintegrasikan faktor cuaca (suhu rata-rata, curah hujan, kelembapan udara) dan kondisi agronomis (luas panen) sebagai input utama. Model *ANN* yang dibangun menunjukkan kinerja yang baik dalam memodelkan hubungan non-linier antara variabel input dan output, serta mampu memberikan hasil prediksi yang akurat berdasarkan evaluasi menggunakan metrik MAE, RMSE, dan  $R^2$ . Proses preprocessing data yang mencakup pembersihan data, penanganan nilai hilang, normalisasi, serta seleksi fitur, telah berhasil meningkatkan kualitas dataset yang digunakan. Arsitektur jaringan *ANN* yang terdiri dari beberapa hidden layer dengan jumlah *neuron* yang disesuaikan serta fungsi aktivasi ReLU menunjukkan performa optimal setelah dilakukan *hyperparameter tuning* menggunakan metode *Grid Search*. *ANN* terbukti unggul dalam memprediksi hasil panen padi berbasis faktor cuaca dan luas panen.

2. Namun, tingkat akurasi ini sangat dipengaruhi kualitas data. Tanpa data cuaca yang akurat dari BMKG dan data pertanian yang konsisten dari BPS, performa model dapat menurun drastis. Berdasarkan analisis terhadap data historis diperoleh bahwa variasi cuaca dan kondisi tanah juga sangat memengaruhi hasil panen padi. Oleh karena itu, pemanfaatan teknologi *ANN* dalam sistem prediksi hasil panen terbukti efektif dan adaptif terhadap dinamika lingkungan.

Secara keseluruhan, penelitian ini menunjukkan bahwa *ANN* merupakan pendekatan yang andal dalam prediksi hasil panen padi, terutama jika dikombinasikan dengan data cuaca dan kondisi tanah yang relevan dan akurat. Hasil penelitian ini diharapkan dapat menjadi kontribusi nyata dalam mendukung sistem pertanian cerdas (*smart agriculture*) serta meningkatkan ketahanan pangan nasional melalui perencanaan produksi yang lebih berbasis data.

## 5.2. Saran

Peneliti menyadari bahwa penelitian yang telah dilakukan terdapat kekurangan yang harus diperbaiki melalui studi-penelitian lanjutan guna mencapai hasil yang lebih optimal. Berikut adalah saran bagi penelitian yang mendatang:

1. Mengintegrasikan variabel tambahan seperti jenis varietas, penggunaan pupuk, dan serangan hama.
2. Menerapkan model *ANN* pada skala lebih luas (provinsi/nasional) untuk melihat generalisasi.
3. Mengombinasikan *ANN* dengan pendekatan *Explainable AI* (XAI) agar hasil prediksi lebih mudah dipahami pemangku kepentingan.
4. Mengembangkan sistem prediksi berbasis *IoT* yang *real-time* agar manfaatnya langsung dirasakan petani di lapangan.

## DAFTAR PUSTAKA

- Abdolrasol, Maher G. M., S. M. Suhail Hussain, Taha Selim Ustun, Mahidur R. Sarker, Mahammad A. HANNan, Ramizi Mohamed, Jamal Abd Ali, Saad Mekhilef, and Abdalrhman Milad. 2021. "Artificial Neural Networks Based Optimization Techniques: A Review." *Electronics* 10(21):2689. doi:10.3390/electronics10212689.
- Ağbulut, Ümit, Ali Etem Gürel, and Yunus Biçen. 2021. "Prediction of Daily Global Solar Radiation Using Different Machine Learning Algorithms: Evaluation and Comparison." *Renewable and Sustainable Energy Reviews* 135:110114. doi:10.1016/j.rser.2020.110114.
- Albahar, Marwan. 2023. "A Survey on Deep Learning and Its Impact on Agriculture: Challenges and Opportunities." *Agriculture (Switzerland)* 13(3). doi:10.3390/agriculture13030540.
- Altunöz, Utku. 2024. "Prediction of Banking Credit Risk Using Logistic Regression and .2The Artificial Neural Network Models: A Case Study of English Banks." *Journal of Social Research and Behavioral Sciences* 10(21):862–87. doi:10.52096/jsrbs.10.21.32.
- Amelia, Poppy, Hatini Diningsih, Siti Nurmaelani, Siti Nurhaeni, JumiYanti JumiYanti, and Yeni Budiawati. 2025. "Artificial Intelligence Untuk Optimasi Efisiensi Dan Produktivitas Hasil Pertanian: Kajian Literatur." *Jurnal Ilmu Pertanian Tirtayasa* 7(1). doi:10.33512/jipt.v7i1.33403.
- Baral, Seshadri, Asis Kumar Tripathy, and Pritiranjana Bijayasingh. 2011. "Yield Prediction Using Artificial Neural Networks." Pp. 315–17 in.
- Bhimavarapu, Usharani, Gopi Battineni, and Nalini Chintalapudi. 2023. "Improved Optimization Algorithm in LSTM to Predict Crop Yield." *Computers* 12(1):10. doi:10.3390/computers12010010.
- Bischi, Bernd, Martin Binder, Michel Lang, Tobias Pielok, Jakob Richter, Stefan Coors, Janek Thomas, Theresa Ullmann, Marc Becker, ANNe-Laure Boulesteix, Difan Deng, and Marius Lindauer. 2023. "Hyperparameter Optimization: Foundations, Algorithms, Best Practices, and Open Challenges." *WIREs Data Mining and Knowledge Discovery* 13(2). doi:10.1002/widm.1484.
- Dhamira, Aura, and Irham Irham. 2020. "The Impact of Climatic Factors Towards Rice Production in Indonesia." *Agro Ekonomi* 31(1). doi:10.22146/ae.55153.



- Do, Huu-Thinh, and Ionela Prodan. 2024. "On the Constrained Feedback Linearization Control Based on the MILP Representation of a ReLU-ANN." *IEEE Control Systems Letters* 8:1445–50. doi:10.1109/LCSYS.2024.3411515.
- Eli-Chukwu, N. C. 2019a. "Applications of Artificial Intelligence in Agriculture: A Review." *Engineering, Technology & Applied Science Research* 9(4):4377–83. doi:10.48084/etasr.2756.
- Eli-Chukwu, N. C. 2019b. "Applications of Artificial Intelligence in Agriculture: A Review." *Engineering, Technology & Applied Science Research* 9(4):4377–83. doi:10.48084/etasr.2756.
- Henderi, Henderi. 2021. "Comparison of Min-Max Normalization and Z-Score Normalization in the K-Nearest Neighbor (KNN) Algorithm to Test the Accuracy of Types of Breast Cancer." *IJIIS: International Journal of Informatics and Information Systems* 4(1):13–20. doi:10.47738/ijiis.v4i1.73.
- Herwanto, Heru Wahyu, Anik Nur Handayani, Aji Prasetya Wibawa, Katya Lindi Chandrika, and Kohei Arai. 2021. "Comparison of Min-Max, Z-Score and Decimal Scaling Normalization for Zoning Feature Extraction on Javanese Character Recognition." Pp. 1–3 in *2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*. IEEE.
- Hodson, Timothy O. 2022. "Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not." *Geoscientific Model Development* 15(14):5481–87. doi:10.5194/gmd-15-5481-2022.
- Irvanizam, Irvanizam, Ivan Horatius, and Hizir Sofyan. 2023. "Applying Artificial Neural Network Based on Backpropagation Method for Indonesian Sign Language Recognition." *International Journal of Computing and Digital Systems* 14(1):975–85. doi:10.12785/ijcds/140176.
- Jiang, Xia, and Chuhan Xu. 2022. "Deep Learning and Machine Learning with Grid Search to Predict Later Occurrence of Breast Cancer Metastasis Using Clinical Data." *Journal of Clinical Medicine* 11(19):5772. doi:10.3390/jcm11195772.
- Joshi, Amit, Jitendra Sasumana, Nillohit Mitra Ray, and Vikas Kaushik. 2021. "Neural Network Analysis." Pp. 351–64 in *Advances in Bioinformatics*. Singapore: Springer Singapore.
- Karunasingha, Dulakshi Santhusitha Kumari. 2022a. "Root Mean Square Error or Mean Absolute Error? Use Their Ratio as Well." *Information Sciences* 585:609–29. doi:10.1016/j.ins.2021.11.036.

- Karunasingha, Dulakshi Santhusitha Kumari. 2022b. "Root Mean Square Error or Mean Absolute Error? Use Their Ratio as Well." *Information Sciences* 585:609–29. doi:10.1016/j.ins.2021.11.036.
- Khairunniza Bejo, Siti, Samihah Mustaffha, Siti Khairunniza-Bejo, Wan Ishak, and Wan Ismail. 2014. "Application of *Artificial Neural Network* in Predicting Crop Yield: A Review." *Journal of Food Science and Engineering* 4(January):1–9.
- Khimzaton, Noor. 2021. "Akhlak Dan Pembentukan Karakter Seseorang." *Jurnal Syntax Transformation* 2(10):1465–79. doi:10.46799/jst.v2i10.433.
- Kühl, Niklas, Max Schemmer, Marc Goutier, and Gerhard Satzger. 2022. "Artificial Intelligence and Machine Learning." *Electronic Markets* 32(4):2235–44. doi:10.1007/s12525-022-00598-0.
- Kumar, Ashish, Abeer Alsadoon, P. W. C. Prasad, Salma Abdullah, Tarik A. Rashid, Duong Thu Hang Pham, and Tran Quoc Vinh Nguyen. 2022. "Generative Adversarial Network (GAN) and Enhanced Root Mean Square Error (ERMSE): Deep Learning for Stock Price Movement Prediction." *Multimedia Tools and Applications* 81(3):3995–4013. doi:10.1007/s11042-021-11670-w.
- Larwuy, Lennox. 2024. "Optimasi Parameter *Artificial Neural Network* (ANN) Menggunakan Particle Swarm Optimization (PSO) Untuk Pengkategorian Nasabah Bank." *Jurnal Matematika Komputasi Dan Statistika* 3(3):506–11. doi:10.33772/jmks.v3i3.60.
- Lawson, Nuanpan. 2023. "New Imputation Method for Estimating Population Mean in the Presence of Missing Data." *Lobachevskii Journal of Mathematics* 44(9):3740–48. doi:10.1134/S1995080223090202.
- Liao, Lizhi, Heng Li, Weiyi Shang, and Lei Ma. 2022. "An Empirical Study of the Impact of Hyperparameter Tuning and Model Optimization on the Performance Properties of Deep Neural Networks." *ACM Transactions on Software Engineering and Methodology* 31(3):1–40. doi:10.1145/3506695.
- Lobell, David B., Wolfram Schlenker, and Justin Costa-Roberts. 2011. "Climate Trends and Global Crop Production Since 1980." *Science* 333(6042):616–20. doi:10.1126/science.1204531.
- Mich, Luisa. 2020. "Artificial Intelligence and Machine Learning." Pp. 1–21 in *Handbook of e-Tourism*. Cham: Springer International Publishing.
- Montesinos López, Osval Antonio, Abelardo Montesinos López, and Jose Crossa. 2022. "Fundamentals of *Artificial Neural Networks* and Deep Learning." Pp. 379–425 in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Cham: Springer International Publishing.

- Nasihin, Sirajun, Stit Palapa, and Nusantara Lombok. 2021. *MENGHAYATI MUKJIZAT ILAHI (Fakta Ilmiah Kemukjizatan Al-Qur'an Dan Sunnah Pada Tumbuhan)*. Vol. 3. <https://ejournal.stitpn.ac.id/index.php/pandawa>.
- Park, Susan. 2020. "World Bank."
- Permana, Inggih, and Febi Nur Salisah Salisah. 2022. "Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma *Backpropagation*." *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)* 2(1):67–72. doi:10.57152/ijirse.v2i1.311.
- Plevris, V., G. Solorzano, N. Bakas, and M. Ben Seghier. 2022. "Investigation of Performance Metrics in Regression Analysis and Machine Learning-Based Prediction Models." in *8th European Congress on Computational Methods in Applied Sciences and Engineering*. CIMNE.
- Qi, Jun, Jun Du, Sabato Marco Siniscalchi, Xiaoli Ma, and Chin-Hui Lee. 2020. "On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression." *IEEE Signal Processing Letters* 27:1485–89. doi:10.1109/LSP.2020.3016837.
- Qodri, Zamzam, and Ahmad Zaidanil Kamil. 2023. "KONTEKSTUALISASI ESKATOLOGIS DI ERA KONTEMPORER." *TAJDID: Jurnal Ilmu Ushuluddin* 22(2):386–426. doi:10.30631/tjd.v22i2.361.
- Ramadhan, Muhammad Rifki, Septiawardi Septiawardi, and Beko Hendro. 2024. "Mitigasi Krisis Pangan Dalam Al-Qur'an (Perspektif Kitab Tafsir Al-Qurtubi)." *Ikhtisar: Jurnal Pengetahuan Islam* 4(2):653. doi:10.55062/IJPI.2024.v4i2/670/5.
- Ratna Sari Dewi, Dea Ananda, Devia Pratiwi, Novi Aulia Safina, and Vira Septria. 2025. "Integrasi Nilai Keislaman Dalam Budidaya Dan Pemasaran Tanaman Hias: Studi Kasus Pada Ganda Nursery." *Maeswara : Jurnal Riset Ilmu Manajemen Dan Kewirausahaan* 3(3):97–112. doi:10.61132/maeswara.v3i3.1801.
- Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?" Pp. 1135–44 in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM.
- Rights, Jason D., and Sonya K. Sterba. 2020. "New Recommendations on the Use of R-Squared Differences in Multilevel Model Comparisons." *Multivariate Behavioral Research* 55(4):568–99. doi:10.1080/00273171.2019.1660605.

- Rosa, João P. S., Daniel J. D. Guerra, Nuno C. G. Horta, Ricardo M. F. Martins, and Nuno C. C. Lourenço. 2020a. "Overview of *Artificial Neural Networks*." Pp. 21–44 in.
- Rosa, João P. S., Daniel J. D. Guerra, Nuno C. G. Horta, Ricardo M. F. Martins, and Nuno C. C. Lourenço. 2020b. "Overview of *Artificial Neural Networks*." Pp. 21–44 in.
- Salmerón Gómez, Román, Ainara Rodríguez Sánchez, Catalina García García, and José García Pérez. 2020. "The VIF and MSE in Raise Regression." *Mathematics* 8(4):605. doi:10.3390/math8040605.
- Sarker, Iqbal H. 2021. "Machine Learning: Algorithms, Real-World Applications and Research Directions." *SN Computer Science* 2(3):160. doi:10.1007/s42979-021-00592-x.
- Saud, Shah, Depeng Wang, Shah Fahad, Hesham F. Alharby, Atif A. Bamagoos, Ali Mjrashi, Nadiyah M. AlabdAllah SWT , Saleha S. AlZahrani, Hamada AbdElgawad, Muhammad Adnan, R. Z. Sayyed, Shafaqat Ali, and Shah Hassan. 2022. "Comprehensive Impacts of Climate Change on Rice Production and Adaptive Strategies in China." *Frontiers in Microbiology* 13(June):1–12. doi:10.3389/fmicb.2022.926059.
- Shahhosseini, Mohsen, Guiping Hu, and Hieu Pham. 2022. "Optimizing Ensemble Weights and Hyperparameters of Machine Learning Models for Regression Problems." *Machine Learning with Applications* 7:100251. doi:10.1016/j.mlwa.2022.100251.
- Singh, Dalwinder, and Birmohan Singh. 2020. "Investigating the Impact of Data Normalization on Classification Performance." *Applied Soft Computing* 97:105524. doi:10.1016/j.asoc.2019.105524.
- Singh, Dalwinder, and Birmohan Singh. 2022. "Feature Wise Normalization: An Effective Way of Normalizing Data." *Pattern Recognition* 122:108307. doi:10.1016/j.patcog.2021.108307.
- Sudin, Sukhairi, Azizi Naim Abdul Aziz, Fathinul Syahir Ahmad Saad, Nurul Syahirah Khalid, and Ismail Ishaq Ibrahim. 2021. "Cycling Performance Prediction Based on Cadence Analysis by Using Multiple Regression." *Journal of Physics: Conference Series* 2107(1):012058. doi:10.1088/1742-6596/2107/1/012058.
- Tyagi, Kanishka, Chinmay Rane, Harshvardhan, and Michael Manry. 2022. "Regression Analysis." Pp. 53–63 in *Artificial Intelligence and Machine Learning for EDGE Computing*. Elsevier.

Urenda, Julio C., and Vladik Kreinovich. 2022. "Why Rectified Linear Activation Functions? Why Max-Pooling? A Possible Explanation." Pp. 459–63 in.

Yang, Li, and AbdAllah SWT Shami. 2020. "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice." *Neurocomputing* 415:295–316. doi:10.1016/j.neucom.2020.07.061.

Zhang, Ze, Xinsheng Chen, Xianyan Qin, Chao Xu, and Xingfu Yan. 2023. "Effects of Soil PH on the Growth and Cadmium Accumulation in Polygonum Hydropiper (L.) in Low and Moderately Cadmium-Contaminated Paddy Soil." *Land* 12(3). doi:10.3390/land12030652.